X3-2 Sub-Committee
on
Character Sets & Data Formats
of the
American Standards Association
X3 Sectional Committee
Sponsored by
O.E.M.I.

March 20, 1961

MINUTES

#6 MEETING OF MARCH 8th & 9th, 1961

SUMMARY OF MINUTES

A large portion of the meeting was devoted to discussion and study of the problems associated with a standard six level, 64 character set. Capt. Luebbert gave a lecture outlining the results of his study of this subject. These deliberations resulted in the identification of two proposed character sets which are to be studied by the membership with the aim that a recommended character set can be identified at the April meeting.

Proposed standards for perforated paper tape have been submitted to the Committee by other organizations. It was decided to table discussion on these items until that time at which the committee undertakes a study of formats and media.

C E Macon

C. E. Macon
Secretary

CEM: aks

NMAH
3/6/1

X3-2 Sub-Committee
on
Character Sets & Data Formats
of the
American Standards Association
X3 Sectional Committee
Sponsored by
O.E.M.I.

March 20, 1961

MINUTES

MEETING OF MARCH 8th & 9th, 1961

1. ATTENDANCE & MEMBERSHIP

1.1 Members Present

John Auwaerter - Teletype Corporation

Leon Bloom - NCR

Theodore H. Bonn - Remington Rand

L. L. Griffin - Navy Department

I. C. Liggett - IBM

C. E. Macon - Burroughs Corporation

Thomas McNamara - Minneapolis Honeywell (Alternate)

Roy W. Reach - Minneapolis Honeywell

H. J. Smith, Jr. - IBM

H. Spielman - RCA (Alternate)

H. Tholstrup - Friden Incorporated

Allen L. Whitman - Bell Telephone Laboratories

1.2 Members Absent

Lawrence J. Schoenberg - Monroe Calculating Machine

B. B. Weber - Addressograph-Multigraph Corp.

1.3 Observers Present

R. W. Bemer

Robert Gryb

W. M. Kahn

W. F. Luebbert

1.4 Resignation

Mr. John H. Craig has voluntarily resigned

1.4 Membership Modification

1.4.1 Notification of Automatic Resignation

Letters had been sent to Mr. Weber and Mr. Schoenberg informing them of the Membership and Participation Rules adopted at the January 11th and 12th meeting of the committee.

Mr. Weber's letter of 2/28/61 was read in which he acknowledged receipt of the letter from the committee and, as a result of studying the rules, submitted his resignation. Mr. Weber requested that he be allowed to attend as an observer. The committee approved Mr. Weber's resignation and request for observer status.

Since Mr. Schoenberg had not replied to the original letter the secretary was requested to inform him of the Membership and Participation Rules by registered mail.

1.4.2 User List

Mr. Auwaerter will submit to the chairman a list of the principle communication user groups by the next meeting.

There are data processing user groups of a general nature in the process of formation (AFIPS & JUG). The chairman will investigate this matter and report at the next meeting.

1.4.3 It was noted that ATA had been invited to join the committee but that no reply had been received as yet.

2. Minutes

The Minutes of the February 9th and 10th meeting were read and approved.

3. Documents

3.1 Documents Distributed

3.1.1 E.I.A. Tentative Standards Proposal for Eleven Sixteenths Inch Wide Five-Track Perforated Paper Tape.

3.1.2 AIEE Standard for Perforated Tape (Preliminary).

3.1.3 Report "On the Design of Extended Character Sets"
by R. W. Bemer

3.1.4 Report "Design of an Improved Transmission/Data Processing Code"
by Bemer, Smith & Williams

NOTE: Replacement copies of the documents distributed to date can be obtained from the Secretary of the Committee.

3.2 The Secretary was directed to obtain for the Committee copies of the report: "Information Handling and Processing In Large Communication Systems" by W.F. Luebbert.

4. Perforated Paper Tape Standards

The subject of the EIA Standards Proposal on One-Inch Perforated Paper Tape distributed at the February meeting was brought forward. The documents listed in 3.1.1 and 3.1.2 were distributed. After some discussion the following motion was approved by the Committee:

A. The documents on Perforated Paper Tape received by the Committee be acknowledged as input information toward that time when the Committee shall consider the subject of Perforated Paper Tape.

B. The subject of standards for perforated paper tape be tabled until that point at which the Committee considers media and format.

5. Six Bit Code Discussion

A six bit code was presented (refer to document under 3.1.4 above). The general subject was discussed at some length and Capt. Leubbert presented a lecture based on his studies of the problem. In addition to the points listed under Section 6 of the February Minutes the following points were brought forward:

5.1 From a communication's point of view it is desirable that codes associated with "double purpose" keys differ in only one binary position for a given key.

5.2 Summary of Discussion on Collating Sequence.

5.2.1 The collating sequence of a code is of primary importance in sorting and forms the basis for the structure of files. In sorting and establishing files all or only a portion of the available character set may be used depending on the Sort Key adopted.

5.2.2 Sort Keys

The basic keys believed to be in use are:

a) Numeric
b) Alpha
c) Alpha & Numeric
c) Alpha & Numeric & Special

5.2.3 If either a numeric only or an alpha only sort key is used then the sequence of the numeric sub-set or the sequence of the alpha sub-set is of prime importance, but the relationship of the numeric sub-set to the alpha is not. In addition, if an alpha-numeric sort key is used and the fields are defined and purely alpha or numeric the same statement holds.

In short, the exact collating sequence of the total character set is of importance if and only if the sort key includes mixed fields of the total character set.

5.2.4 Importance of Collating Sequence

It was noted that from a purely logical point of view virtually any well ordered and defined sequence is satisfactory if new files are being sorted and established. In the case of established files, however, changing the sort key could require a considerable re-sorting task which would be costly. It is believed that for one large group of data processing users that about 60% of this group are using either pure alpha or pure numeric keys and hence will not be effected unless the normal sequence of these sub-sets is disturbed. Of the remaining 40% which use alpha and numeric and limited specials, the majority employ a special, alpha, numeric sequence.

5.3 General Six Bit Set

In considering a General Six Level Set for communications and data processing, it seems desirable the five level set derived from collapsing the six level set provide for something less than 64 characters. Since 6 special function characters are required (i.e., Master space, blank, carriage return, line feed, escape and delete) and, as presently envisioned, the corresponding 6 characters in the six level code will be lost in the "collapsing" process, it appears that 52 characters are available for general six level interchange of information.

6. First Draft Proposals - Homework

Two six level codes were identified (see Table 1 and 2) which seemed to accomplish most of the objectives. These are to be studied so that a single proposed six level code can be identified at the next meeting.

7. Communications

In his letter of March 3, 1961, Mr. Liggett acknowledged establishment of Mr. Robert J. Rossheim as the coordinator between X3-2 and X3-6.

8. Next Meeting

The next meeting will be held April 12th and 13th. The exact time and place will be stipulated in the forthcoming Agenda.

C E M

C. E. Macon
Secretary

CEM:aks

	00	01	02	03
00	Master Space	0	<	M
01	ƀ	1	=	N
02	CR	2	>	O
03	LF	3	@	P
04	#	4	A	Q
05	%	5	B	R
06	:	6	C	S
07	;	7	D	T
10	(	8	E	U
11	)	9	F	V
12	/	$	G	W
13	,	*	H	X
14	—	?	I	Y
15	+	!	J	Z
16	.	UC	K	Escape
17	;	LC	L	Delete

TABLE 1

00	01	02	03
Master Space 🔔	0	①	O
ƀ	1	①	P
"	2	A	Q
#	3	B	R
$	4	C	S
%	5	D	T
,	6	E	U
&	7	F	V
(	8	G	W
)	9	H	X
?	/	I	Y
WRU	*	J	Z
—	:	K	CR
+	;	L	LF
.	UC	M	Escape
;	LC	N	Delete

① To be selected

TABLE 2

First draft proposals - X3.2 61 Mar 8,9
to X3.2

BSW?

Memorandum

FOR ALL INTERNAL COMPANY CORRESPONDENCE

DATE March 14, 1961

TO J. E. Smith, W. W. Finke

FROM R. W. Reach

SUBJECT X3-2 Subcommittee Meeting

COPIES TO:
H. A. Antman
B. K. Betz
R. M. Bloch
R. F. Clippinger
C. F. Dubay
J. J. Backus
T. J. McNamara

The ASA X3-2 Subcommittee on Character Sets and Data Format met in New York on Wednesday and Thursday, March 8 and 9.

Mr. Bruce Weber of Addressograph-Multigraph has resigned from the committee and Mr. Larry Schoenberg of Monroe is being dropped from the committee due to lack of attendance. The membership now stands at eleven and an attempt to get new members from several users' groups--Air Transport Association, AFIPS, JUG, etc.--is being made.

The EIA Proposed Standard for one inch paper tape and a competitive Proposed Standard from AIEE were submitted to X3-2 for approval. Further action by X3-2 was tabled until the committee (X3-2) arrives at a point where it can consider physical considerations some time in the future.

A compromise proposed 6-bit code was submitted by IBM in an attempt to bring agreement on the code selected by X3-2. Captain Luebbert of the U. S. Army made a presentation essentially in support of a slightly modified Fieldata Code. As in the previous meetings, the delegates from AT&T and Teletype made very vocal demands about the code, based on the savings to them of a relay contact or coil. However, the committee now seems to be considering the long-term effects of the code on a more generalized basis. After considerable debate, the code attached to this memorandum was agreed on as a first approach, reaching almost unanimous agreement. It will be circulated for comments within all companies represented on X3-2.

The next meeting is scheduled for April 12 and 13 in New York, during which comments regarding the proposed 6-bit code set, possible 4-bit subset, 7- and 8-bit codes, and collating sequence will be discussed.

R. Reach

RR:RJ

CBI 64/1/1

First Proposal for

Standard Code Set by EIA TR-[illegible]

Octal	Character	Octal	Character	Octal	Character	Octal	Character
00	Master Space	20	0	40	(to be assigned)	[illegible]	[illegible]
01	blank	21	1	41	(to be assigned)	[illegible]	[illegible]
02	Carriage Return	22	2	42	(to be assigned)	[illegible]	[illegible]
03	Line Feed	23	3	43	(to be assigned)	[illegible]	[illegible]
04	/	24	4	[illegible]	[illegible]	[illegible]	[illegible]
05	&	25	5	[illegible]	[illegible]	[illegible]	[illegible]
06	[illegible]	26	6	[illegible]	[illegible]	[illegible]	[illegible]
07	[illegible]	27	7	[illegible]	[illegible]	[illegible]	[illegible]
10	[illegible]	30	8	[illegible]	[illegible]	[illegible]	[illegible]
11	[illegible]	31	9	[illegible]	[illegible]	[illegible]	[illegible]
12	[illegible]	32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	[illegible]	33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	-	34	(to be assigned)	[illegible]	[illegible]	[illegible]	[illegible]
15	[illegible]	35	(to be assigned)	[illegible]	[illegible]	[illegible]	[illegible]
16	[illegible]	36	Upper Case	[illegible]	[illegible]	[illegible]	[illegible]
17	[illegible]	37	Lower Case	[illegible]	[illegible]	[illegible]	[illegible]

The following [illegible] are features of this code:

1. The alphabetics are in a 6-bit subset.

2. [illegible]

3. [illegible]

[illegible]

4. [illegible]

[illegible]

[illegible]

5. [illegible]

[illegible]

Mr. Tom [illegible], the Honeywell delegate to the EIA TR 27.6 and the alternate to X3-2, also attended.

First Proposal for
Standard Code Set by ASA X3-2

Octal	*Character*	*Octal*	*Character*	*Octal*	*Character*	*Octal*	*Character*
00	Master Space	20	0	40	(to be assigned)	60	M
01	blank	21	1	41	(to be assigned)	61	N
02	Carriage Return	22	2	42	(to be assigned)	62	O
03	Line Feed	23	~~4~~ 3	43	(to be assigned)	63	P
04	#	24	4	44	A	64	Q
05	%	25	5	45	B	65	R
06	:	26	6	46	C	66	S
07	;	27	7	47	D	67	T
10	(	30	8	50	E	70	U
11	)	31	9	51	F	71	V
12	/	32	$	52	G	72	W
13	'	33	*	53	H	73	X
14	-	34	(to be assigned)	54	I	74	Y
15	+	35	(to be assigned)	55	J	75	Z
16	.	36	Upper Case	56	K	76	Escape
17	,	37	Lower Case	57	L	77	Delete

The following points are features of this code:

1) The Alphabetics are in a 5-bit subset
2) The format effectors (MS, Blank, CR, and LF) collate first and are easy to logically recognize
3) Master space precedes all characters so that it is sure to collate first
4) All punctuation marks collate before numerics and alphabetics
5) Two symbols ($ and *) unique to the U.S. occupy codes 32 and 33 following the numeric 0-9 so that compatibility with a British duo-decimal code is achieved.
6) By collapsing the low order six-bit subset into a four-bit subset, codes 20-33 and ~~34-37~~ 14-17 become a convenient numeric subset
7) If a numeric 4-bit subset or an alphanumeric 5-bit subset is desired, certain pending positions in each set are available for control or format effectors (i.e. the function assigned octal 00-03) can be assigned to and the similarity between octal 36-37 and octal 76-77 allow them to be interchangeable.
8) Delete is all ones (octal 77) to allow paper tape punch-out
9) Six code positions are available for special applications

Mr. Tom McNamara, the Honeywell delegate to the EXA TR 27.6 and the alternate to X3-2, also attended.

CERTIFICATION OF ALGORITHM 20
REAL EXPONENTIAL INTEGRAL (S. Peavy, *Comm. ACM*, Oct. 1960)

William J. Alexander* and Henry C. Thacher, Jr.*
Argonne National Laboratory, Argonne, Illinois

Expint (x) was programmed for the LGP-30 computer, using both a 7S floating-point compiler (ACT III) and an 8S floating-point interpretive code (24.2). Constants given to more than 7S (or to 8S for the 24.2 program) were rounded to 7S (or 8S).

After changing the constant .005519968 to .05519968, both programs gave acceptable accuracy over the range tested.

The 8S (24.2) program was compared with the 9D values given for $-E_1(-x)$ in Mathematical Tables Project, *Tables of Sine, Cosine, and Exponential Integrals, Volume* II (1940), for the set of values $x = 0.1(0.1)1.0(1.0)10.0$. The largest discrepancy found was -16×10^{-8} for $x = 0.1$. For x greater than 1, all values tested were good to 8S.

For computing real values of the exponential integral, this algorithm is much faster than EKZ (Algorithm 13). For $x < 1$, the ratio of speeds was of the order of 20.

---

* Work supported by the U.S. Atomic Energy Commission.

CERTIFICATION OF ALGORITHM 43
CROUT II (Henry C. Thacher, Jr., *Comm. ACM*, 1960)

Henry C. Thacher, Jr.*
Argonne National Laboratory, Argonne, Illinois

CROUT II was coded by hand for the Royal Precision LGP-30 computer, using a 28-bit mantisa floating point interpretive system (24.2 modified).

The program was tested against the linear system:

$$A = \begin{pmatrix} 12.1719 & 27.3941 & 1.9827 & 7.3757 \\ 8.1163 & 23.3385 & 9.8397 & 4.9474 \\ 3.0706 & 13.5434 & 15.5973 & 7.5172 \\ 3.0581 & 3.1510 & 6.9841 & 13.1984 \end{pmatrix} \qquad b = \begin{pmatrix} 6.6355 \\ 6.1304 \\ 4.6921 \\ 2.5393 \end{pmatrix}$$

with the following results:

$$A' = \begin{pmatrix} 12.171900 & 27.394100 & 1.9827000 & 7.3756999 \\ 0.25226957 & 6.6327021 & 15.097125 & 5.6565352 \\ 0.25124262 & -0.56260107 & 14.979620 & 14.527683 \\ 0.66680633 & 0.76468695 & -0.20207132 & -1.3606142 \end{pmatrix}$$

$$b' = \begin{pmatrix} 6.6354999 \\ 3.0181653 \\ 2.5702026 \\ -0.082780734 \end{pmatrix} \qquad \text{pivot} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix} \qquad y = \begin{pmatrix} 0.15929120 \\ 0.14691771 \\ 0.11257482 \\ 0.060840712 \end{pmatrix}$$

det = −1645.4499. All elements of $Ab - y$ were less than 10^{-7} in magnitude. Identical results were obtained with the same b, and repeat **true**. With the same b and the last row vector of A replaced by (19.1927, 33.4409, 25.1298, 5.2811), i.e. $A\,4, j = A\,1\,j, + 2A\,2, j - 3A\,3, j$, the results were:

$\det = 0.10924352 \times 10^{-3}$,

$y = (0.29214425 \times 10^{8}, -0.12131172 \times 10^{8}, 0.72411923 \times 10^{7}, -0.51018392 \times 10^{7})$

Failure to recognize this singular matrix is due to roundoff, either in the data input or in the calculation.

---

* Work supported by the U.S. Atomic Energy Commission.

Standards

Further Survey of Punched Card Codes

H. McG. Ross, *Ferranti Ltd., London*

The valuable "Survey of Punched Card Codes" prepared by Smith and Williams (*Comm. ACM 3*, Dec. 1960, 638) unfortunately omits the card codes of European equipment, other than IBM. These are presented in the table on page 181. This information has been extracted from a Ferranti publication, "Collected Information on Punched Card Codes" (List CS 266) and has been set out in much the same way as the table by Smith and Williams.

A valuable step forward has been made by the British Standards Institution in publishing Standard 3174, "Alpha-Numeric Punching Codes for Data Processing Cards". As well as decimal numbers and letters, this Standard also gives single-column punching codes for pence, for shillings, for months, and for days within the month, etc.

In the table, the card rows are identified by 12, 11, 0, 1, · · · , 9, from top to bottom; it should be noted, however, that modern practice in Britain prefers the top row to be numbered 10. Where nothing is punched in a zone or position the symbol b̸ is used. The abbreviation Sp is used for the space obtained in a printer from an entirely blank column. Letter O is shown with a dot in it, to distinguish it from zero.

The table gives examples of the treatment of days, months, pence, etc., particularly when a single column is used, often with two holes in it. However, even within one type of code, 10 and 11 may be reversed.

The "old Hollerith" 4-zone code, which is widely used, is designed to "cycle" within the zones; this brings the letters into alphabetical order.

The asterisks refer to the Bull codes in which rows 7, 8, and 9 are used to punch the zones, in place of 12, 11, and 0 respectively. Another special optional feature is the "mechanical zero" punched in row 12; this is for left-hand zeros, and is treated as zero by the accounting machine but does not print.

→

v4 #4
April 1961

Survey of Punched Card Codes—European Equipment

	b	12	11	0	1	2	3	4	5	6	7	8	9
	b												
British Standard	Sp	10	11	0	1	2	3	4	5	6	7	8	9
I.C.T. (New Hollerith) 5-zone	Sp	10	11	0	1	2	3	4	5	6	7	8	9
I.C.T. (New Hollerith) 4-zone	Sp	10	11	0	1	2	3	4	5	6	7	8	9
I.C.T. (Revised Hollerith) 3-zone	Sp	10	11	0 or Θ	1 or I	2	3	4	5	6 or G	7	8 or S	9
B.T.M. (Old Hollerith) 4-zone	Sp	11	10	0	1	2	3	4	5	6	7	8	9
Powers 40-character	Sp	10	11	0	1	2	3	4	5	6	7	8	9
Powers 32-character	Sp	10	11	0 or O	1 or I	2	3	4	5	6 or G	7	8 or S	9
Powers 29-character	Sp	11	10	1 or I	2 or Z	3	4	5	6 or G	7	8 or S	9	
Bull* Standard	Sp	11 or 0	10 or .	0 or Θ	1 or I	2	3	4	5	6	7	8	9
Bull* 300 Series	Sp	11	10	0	1	2	3	4	5	6	7	8	9

	12 or 7*											
	12	11	0	1	2	3	4	5	6	7	8	9
British Standard				A	B	C	D	E	F	G	H	I
I.C.T. (New Hollerith) 5-zone				A	B	C	D	E	F	G	H	I
I.C.T. (New Hollerith) 4-zone				A	B	C	D	E	F	G	H	I
I.C.T. (Revised Hollerith) 3-zone	Z		N	P	Q	R	T	U	V	W	X	Y
B.T.M. (Old Hollerith) 4-zone				A	D	G	J	M	P	S	V	Y
Powers 40-character				A	B	C	D	E	F	G	H	I
Powers 32-character	A		C	D	E	F	H	J	K	L	M	N
Powers 29-character	A		B	C	D	E	F	H	J	K or X	L	
Bull* Standard		A	B	C	D	E	F	G	H			
Bull* 300 Series	Θ	A	B	C	D	E	F	G	H			

	11 or 8*											
	12	11	0	1	2	3	4	5	6	7	8	9
British Standard				J	K	L	M	N	O	P	Q	R
I.C.T. (New Hollerith) 5-zone				J	K	L	M	N	Θ	P	Q	P
I.C.T. (New Hollerith) 4-zone				J	K	L	M	N	Θ	P	Q	R
I.C.T. (Revised Hollerith) 3-zone		A	B	C	D	E	F	H	J	K	L	M
B.T.M. (Old Hollerith) 4-zone				B	E	H	K	N	Q	T	W	Z
Powers 40-character				J	K	L	M	N	O	P	Q	R
Powers 32-character		B	P	Q	R	T	U	V	W	X	Y	Z
Powers 29-character		O or 0	M	N	P	Q	R	T	U or V	W	Y	
Bull* Standard		J	K	L	M	N	P	Q	R			
Bull* 300 Series	I	J	K	L	M	N	P	Q	R			

	0 or 9*											
	12	11	0	1	2	3	4	5	6	7	8	9
British Standard				&	S	T	U	V	W	X	Y	Z
I.C.T. (New Hollerith) 5-zone				&	S	T	U	V	W	X	Y	Z
I.C.T. (New Hollerith) 4-zone				&	S	T	U	V	W	X	Y	Z
I.C.T. (Revised Hollerith) 3-zone												
B.T.M. (Old Hollerith) 4-zone				C	F	I	L	Θ	R	U	X	&
Powers 40-character				S	T	U	V	W	X	Y	Z	&
Powers 32-character												
Powers 29-character												
Bull* Standard		S	T	U	V	W	X	Y	Z			
Bull* 300 Series	.	S	T	U	V	W	X	Y	Z			

	1								9 + 7										9 + 8									
	2	3	4	5	6	7	8	9	b	12	11	0	1	2	3	4	5	6	b	12	11	0	1	2	3	4	5	6
British Standard																												
I.C.T. (New Hollerith) 5-zone	%	¼	–	/	½	.	@	¾																				
I.C.T. (New Hollerith) 4-zone																												
I.C.T. (Revised Hollerith) 3-zone																												
B.T.M. (Old Hollerith) 4-zone																												
Powers 40-character																												
Powers 32-character																												
Powers 29-character																												
Bull Standard																												
Bull 300 Series									*	/	,	'	%	$	£	◇	□	△	–	+	=	(	)	×	:	<	>	≠

	Fractions			Pence			Shillings			Days			Months		
	3	6	9	12	11	0	12	11	0	12	11	0	12	11	0
British Standard				10	11	0	10	20	0	10	20	30	Θ	N	D
I.C.T. (New Hollerith) 5-zone				10	11	0	10	20	0	10	20	30	Θ	N	D
I.C.T. (New Hollerith) 4-zone	¼	½	¾	10	11	0	10	20	0	10	20	30	Θ	N	D
I.C.T. (Revised Hollerith) 3-zone				10	11	0									
B.T.M. (Old Hollerith) 4-zone				11	10	0									
Powers 40-character	¼	½	¾	10	11	0									
Powers 32-character				10	11	0									
Powers 29-character				11	10	0									
Bull Standard				11	10	0									
Bull 300 Series				11	10	0									

H. McG. ROSS, JANUARY 1961

Subcommittee X3-2 Character Sets and Data Formats of ASA X3 Sectional Committee on Data Processing Standards

MINUTES

Sixth Meeting, April 12th and 13th, 1961
ASA Headquarters, New York, New York

SUMMARY

The relationship of the six bit (64 character) coded set to larger (7 and 8 bit) and smaller (4 and 5 bit) sets must be determined. This relationship must be delineated and included in the recommended standard. Accordingly the majority of the meeting was devoted to exploring the content of the various sets as a function of set size and determining the conversional modes between these sets. An interesting structure for a family of coded character sets spanning the range from a 4 bit set through an 8 bit set was identified. These considerations will be pursued at the next meeting scheduled for April 26th and 27th in New York.

NMAH 310/1

April 19, 1961

X3-2 Sub-Committee
on
Character Sets & Data Formats
of the
American Standards Association
X3 Sectional Committee
Sponsored by
O.E.M.I.

MINUTES

MEETING OF APRIL 12th AND 13th

1. MEMBERSHIP & ATTENDANCE

1.1 Members Present

John Auwaerter - Teletype Corp.

Leon Bloom - NCR

Theodore H. Bonn - Remington Rand

L. L. Griffin - Navy Department

I. C. Liggett - IBM

C. E. Macon - Burroughs Corp.

Roy W. Reach - Minneapolis Honeywell

H. J. Smith, Jr. - IBM

H. Tholstrup - Friden Inc.

Allen L. Whitman - Bell Telephone Lab.

1.2 Members Absent

H. Kleinberg - RCA

Lawrence J. Schoenberg - Monroe Calculating Machine

1.3 Observers Present

Robert Gryb - AT&T Company

Capt. W. J. Luebbert - West Point

1.4 Changes in Membership

Mr. Robert Gryb was elected to membership by unanimous vote.

Capt. W. J. Luebbert was elected to membership by unanimous vote.

Mr. Schoenberg was contacted by phone in addition to the two letters previously sent. He will submit a letter by the next meeting clarifying his intent with respect to the Committee activities.

2. REVIEW OF MINUTE DISTRIBUTION

The subject of Minute distribution was briefly discussed. Mr. W.S. Nevison of Recordak and Mr. Rosshein, liaison X3-6 to X3-2, are to receive minutes. These are in addition to the normal distribution.

3. CORRESPONDENCE

3.1 Sent

3.1.1 Letter to Mr. Schoenberg requesting that he clarify his status with respect to the Committee.

3.1.2 Letter to Stanford University requesting 20 copies of W. J. Luebbert's Technical Report entitled "Information Handling & Processing in Large Communication Systems".

3.2 Received

3.2.1 Letter - Bright to Liggett and members: attached "Further Survey of Punched Card Codes" H. McG. Ross, Ferranti Ltd., London.

3.2.2 Letter - Bright - Format for Cover Sheets of Minutes of Meetings.

3.2.3 From Bright - Revision of Fact Sheet on the Data Processing Group.

3.2.4 In reply to an earlier inquiry, Mr. Ross of Ferranti sent a copy of his report. "Considerations in Choosing a Character Code for Computers and Punched Tapes".

4. DOCUMENTS DISTRIBUTED

4.1 Draft report by Smith on the philosphy of structuring a family of coded character sets.

4.2 Draft of the "Report on EIA Survey of Information Processing".

4.3 "An 8 Bit Code Set Structure" by J. F. Auwaerter.

5. USER GROUPS

5.1 JUG

The Chairman reported that he had contacted Mr. Harry Cantrell, President of J.U.G. Mr. Cantrell indicated that J.U.G. would like to cooperate and would be willing to:

a) Supply a representative to X3-2
b) Set up separate task groups to work in certain areas and/or
c) Supply information on request from X3-2

X3-2 will consider which of these courses will be of the most value to both groups.

5.2 A.F.I.P.S.

The Chairman will continue his investigations in this area.

5.3 Communications

Mr. Auwaerter submitted a list of four of the principle user groups in the communication area. (See Attachment A).

Mr. Howard reported briefly on his visit to the recent Air Transport Association meeting. Their work at this point is not directly related to the present considerations of X3-2.

6. X3 MEETING

The Chairman gave a brief report on the recent X3 Sectional Committee meeting. A copy of the Status Report which he presented at this meeting is attached. (See Attachment B).

7. MINUTES

The Minutes of the X3-2 meeting of March 8th and 9th, 1961 were approved.

8. DISCUSSION OF CODED CHARACTER SET FAMILIES

8.1 Mr. Smith presented a discussion of the draft report on this subject (reference 4.1 above)

8.2 In the open discussion which followed the considerations which must be included in determining the structure, content and interrelationships involved in such a family of sets were examined in some detail. It was determined that the scheme presented in the draft report was very sound from several points of view. Some of the points brought out in the discussion were:

8.2.1 Collapsing a Six Bit Code into a Five Bit Code

Some light investigations which have been made on this subject indicate that there is little interest in this feature unless, of course, the collapse is to the 5 Bit Baudot Code (which is highly unlikely if a simple scheme is invoked).

8.2.2 Double Purpose Keys

With regard to the code difference between characters carried on "double purpose keys" it was noted that a 2 bit (as opposed to the previously noted 1 bit) difference is permissible if the difference is constant. (e.g. 00XXXX for the lower case and 11XXXX for the upper case).

8.2.3 Collating Sequence

8.2.3.1 It is desirable that the basic sequence, (that is, the relationship between the alphabetic, numeric, special symbol, and control sub sets) be the same for the 6 bit, 7 bit, and 8 bit sets.

8.2.3.2 Effect on existing files

A light survey conducted by one of the members seemed to indicate the following situation with existing files:

a) 60% of the total employ an alpha or numeric or alpha and numeric sort key in fixed fields. These files will not be effected by a change in basic sequence.

b) 40% of the total employ a mixed alpha and numeric sort key and would be effected by a change in basic sequence.

c) 25% of the total employ a mixed alpha and numeric sort key (and are included in (b) above) with the basic sequence of specials low, then alphabet and numerics high.

8.2.3.3 Location of alphabet

Locating the alphabet in either the first two or last two quadrents of the 6 bit set has the following advantages:

a) The code is easier to handle in "primitive" (i.e. simple) equipment. (e.g. key set equipment, paper tape readers and punches, small printers and the like).

b) The code is more easily expanded to larger sets or contracted to smaller sets.

The disadvantage is the effect noted in 8.2.3.2, above.

8.2.4 7 Bit Set

The point was brought forward that it might be desirable to standardize on a seven bit set instead of a six bit set as the

---

standard for information interchange for communications and data processing. This set would expand to eight bits when parity is added. This point is to be given some study by the members and will be considered at the next meeting.

9. BASIC CRITERIA - 64 CHARACTER SET

A special task group derived a set of basic criteria for consideration by the Committee. Time did not permit discussion of the criteria at this meeting. The task group proposal is attached for study and will be discussed at the next meeting. (See Attachment C).

10. MEETING

The Committee feels that there is an urgent need for a standard character set. Accordingly the following meeting schedule has been established:

April 26th & 27th Regular Meeting

Starting Time: 9:30 A.M. 4/26/61
Place: Engineers Club (3rd Floor)
32 West 40th Street
New York, N. Y.

May 8th (Mon.) & 11th (Thurs.) Informal Meeting

This meeting will consist of those who are attending the Western Joint Computer Conference.

Exact time and place will be announced at the April 26th & 27th meeting.

June 7th & 8th Regular Meeting

Either ASA Headquarters or the Engineers Club, New York City. The exact time and location will be announced.

C E Macon
C. E. Macon
Secretary

CEM:aks

ATTACHMENT A

B6
B4

INDUSTRIAL COMMUNICATIONS ASSOCIATION

Mr. Vincent Murphy
Union Carbide Corporation
30 E. 42nd Street
New York 17, New York

AIR TRANSPORT ASSOCIATION OF AMERICA

Mr. Frank C. White
ANTC Division
Air Transport Association of America
1000 Connecticut Avenue, N. W.
Washington 6, D.C.

MILITARY

Mr. R. S. Boykin, Chairman
Committee on Military Communication System
Technical Standards
U.S. Army Signal Research and Dev. Lab.
Fort Monmouth, New Jersey

ASSOCIATION OF AMERICAN RAILROADS

Mr. A. Grothman
Secretary of Communications & Signal Section
59 E. Van Buren Street
Chicago, Illinois

ATTACHMENT B

STATUS REPORT

ASA X3-2 SUBCOMMITTEE

The following people are presently active on the X3-2 Subcommittee:

I. C. Liggett, Chairman	IBM
C. E. Macon, Secretary	Burroughs
John Auwaerter	Teletype
Leon Bloom	NCR
T. H. Bonn	Remington Rand
L. L. Griffin	DOD (Navy Dept.)
H. Kleinberg	RCA
R. W. Reach	M-H
L. J. Schoenberg	Monroe
H. J. Smith, Jr.	IBM
H. Tholstrup	Friden
A. L. Whitman	Bell Labs.

Captain W. F. Luebbert tentatively joined the group as of March 8. He was instrumental in the design of the Fieldata Code.

The following people have resigned from the committee:

J. H. Craig	AT&T
E. A. Emerson	NCR
R. W. Green	Standard Register
W. S. Nivison	Recordak
E. C. Olofson	Royal McBee
B. B. Weber	A-M

Our original objectives were to establish character content, collating sequence, and the actual bit representation and then iterate over these criteria to a solution. We have tentative recommendations for a character set. As a result of this process, we have developed the following design objectives:

1) Should be similar to existing codes and/or a union of existing codes.

2) Contain same number of bits per character.

3) Unique codes for alphabet, numbers, control, and as many special symbols as possible.

4) Logical structure for expansion and contraction.

5) Provide simple means of categorizing groups of characters.

6) Numbers should be BCD (Binary Coded Decimal).

7) Maintain historic collating sequence.

8) Simple sorting procedure to maintain collating sequence.

9) Contiguous alphabet and contiguous numbers.

10) Expansion for foreign alphabet, and allow for 10 - 11 for Britain.

11) Suitable for data transmission.

12) Must contain data format control and equipment control characters.

13) Provide logical means of using large alternate sets on an exception basis (ESCAPE).

This tentative proposal has about an 80% chance of gaining acceptance by May 1961 within our committee. We believe it represents the best solution to the criteria mentioned with the least compromise.

This character set is thought of as a data processing-communication code. We believe it is organizationally sound, usable in 4, 5, 6, 7, and 8 bit versions and has a probable life of 15 years. Our immediate schedule contains the following:

1) Reach agreement on a final proposal.

2) Work with the Fieldata people and determine their reaction to this new code.

I. C. Liggett 3/27/61

April 19, 1961

ATTACHMENT C

BASIC CRITERIA FOR 64 CHARACTER SET

duction

following criteria were compiled by an informal task group working the evening April 12th. Reference is made to Fieldata, Section 3.6 Criteria for Alpha- meric Code Selection and the criteria set forth in the EIA Tentative Standards roposal for Basic Character Set Code of May 25, 1960. The 10 points set forth in ieldata are included and all but 2 of the points of the EIA proposal. The EIA points not included are:

H) The numerals should be contiguous with the alphabet.

J) The numerics should follow the alphabet.

1) The Set is to facilitate the interchange of information between Systems and between Systems and Associated Equipment.

FD-5 2) The set is to include:

- a) The alphabet (Upper Case)
- b) The numerals 0 - 9
- c) Certain special characters
- d) The minimum control codes required to implement (1) above

EIA-E-C FD-2 3) The codes for all symbols of the set are to include the same number of bits (6).

EIA-E 4) The code set is to be structured such that expansion to larger sets or contraction to smaller sets can be achieved in a simple and logical manner.

EIA-D FD-3 5) The basic set and code shall not contain error control.

EIA-B&A 6) The alphabetic sub set shall be contiguous.

April 19, 1961

3-2)
ia-64 Char. Set

7) The numeric sub set shall be contiguous.

8) The set shall be so organized and coded that the simplest possible test shall be adequate to distinguish and identify the basic alphabetic, numeric and symbolic sub sets.

L 9) The numeric sub set shall be so coded that the least four significant bit shall be the binary coded decimal form of the decimal digit which the code represents.

-9 10) The alphabetic sub set shall be so coded that the code for A will be the smallest binary number and the code for Z the largest binary number within the alphabetic code area. (Of course the code and alphabet are monotonically increasing from A → Z).

EIA-I 11) The code set shall contain an "Escape" character and code.

12) If possible, the set will include the COBOL character set.

EIA-K 13) The code 000000 will be reserved for a space or blank function.

14) The word separator must be "low".

EIA-G 15) Within the set each character and corresponding code will stand by itself and not depend on surrounding characters in a string for interpretation.

16) The binary sequence and the collating sequence of the character set be identical.

17) The all ones code will be reserved for a delete function.

AN 8 BIT CODE SET STRUCTURE

$B_3B_2B_1B_0$	00				01				10				11				B_7 & B_6
	00	01	10	11	00	01	10	11	00	01	10	11	00	01	10	11	B_5 & B_4
0 0 0 0	MS b̄ S	D	A	Z	?	?	C	C	?	?	a	z	?	?	C	C	
⋮																	
1 1 1 1																	

Notation: MS is communications BLANK
Fieldata Master Space - all 0's

b̄ is data BLANK, communications SPACE

S is Symbols

D is Digits

A - Z are upper case Alphas

C is Control

a - z are lower case Alphas

Characteristics:

1. Provides a 4 bit subset in column $B_7B_6B_5B_4$ → 0001
2. Strips to a 5 bit subset in columns $B_7B_6B_5B_4$ → 0000 and 0001
   Master Space and Blank (Space)
   collate low, as special graphics
3. Strips to a 6 bit subset in columns B_7B_6 → 00
4. The 4, 5 and 6 bit subsets have no control characters
5. Provides a 7 bit set in B_7B_6 → 00 and 01
6. Eight bit set provides lower case alphas and more controls
7. Controls all have B_6B_5 → 11
8. Alphas all have B_6B_5 → 01
9. Upper case alphas differ from lower case alphas only in bit B_7
10. Provides room for expansion and "do it yourself" areas

J. F. AUWAERTER

NMAH 310/1

DATA EQUIPMENT BRANCH

MEMORANDUM REPORT

SIGRA/SL-NPE

14 April 1961

MEMORANDUM FOR: Director, Data Processing Facilities Division

SUBJECT: Some Possible Future Changes That Will Not Affect Present Systems and Equipments in the FIELDATA Keyboard and Code, and the Effects of These Changes

PURPOSE: The purpose of this memorandum is to state some changes that may be desirable in the FIELDATA Keyboard and Code. The desirability of these changes will be discussed along with the estimated effects on the FIELDATA system. It is hoped that the suggestions and comments received from persons involved in the FIELDATA program will give rise to a final report of recommended changes.

BACKGROUND: The Minutes of the Sixteenth Meeting of the Third Review of the Committee on Military Communication System Technical Standards, dated 30 January 1961, contained a proposed modification to the FIELDATA code by the Teletype Corporation (Bell System) which they propose using for a new teletypewriter service under consideration by the Bell System.

Since the ADPS Keyboard has certain inconsistencies with communications oriented devices it would be desirable to modify the FIELDATA code and Keyboard if such changes were tolerable. It should be emphasized that these changes do not affect present equipments, systems and items under development.

DISCUSSION OF CHANGES: The following will not be changed:

Master Space
Upper Case
Lower Case
Line Feed
Carriage Return
Space
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
)
+
*
(
"
:
?
,
0
1
2
3
4
5
6
7
8
9
'
;
/
Special

Info. Only

The following are possible changes:

-	changed	to	!
$	"	"	&
!	"	"	_
.	"	"	-
delete	"	"	idle
idle	"	"	delete
>	"	"	#
=	"	"	$
<	"	"	%
Blank/idle	"	"	blank
	"	"	bell
⊕ ⊕	"	"	.

The above changes were motivated by:

(a) An attempt to arrange the keys so that they would be more in line with commercial keyboards.

(b) An attempt to make FIELDATA alphanumeric compatible with a 6 level commercial code (ATT).

(c) An attempt to replace symbols which would be rarely used by possible more useful symbols.

DISCUSSION OF CHANGES:

BLANK/IDLE to BLANK would only involve a change in name. This change would avoid ambiguity (since there is an idle code), and would better describe the code function.

The changes would enable placing _"underline" and - "dash" on the same font, since they now differ only in one bit with respect to each other. Also since "dash" is used for "minus" it would now be in the same case reference to the numerical codes.

The exclamation mark (!) and the Bell are now able to be placed over the 1 and 6.

"Greater than" > , "equal to" =, and "less than" < were deleted and replaced by #, $, and % respectively. These substitutions are possibly more useful.

The ⊕ symbol was deleted and the code assigned to "period". This change enables both "comma" and "period" to each have a separate key for

good

both upper and lower cases. This is identical with commercial practice.

The idle code combination would now represent delete, and the delete code combination would now represent idle. Idle would therefore now be a control character, and delete would now be an alphanumeric character. The major benefit of this change is that the FIELDATA paper tape code can now be slightly modified to enable the conversion process of Standard Form to Paper tape Form (and vice versa) to be realized by a simple inversion of the parity bit. Another advantage is that the FIELDATA "delete" is now a true 6 bit code delete.

By depressing the control code key the keyboard can be used to generate all 64 control characters.

Note: Although there is no backspace key on the proposed keyboard, the keyboard will be capable of generating and functionally responding to the control code for backspace.

Prepared by

Division Keyboard Committee

ALPHANUMERIC

PROPOSED CHARACTER	STANDARD CODE (ODD PARITY)								PAPER TAPE (EVEN PARITY)								FIELDATA	ATT
	P 7	C 6	I_2 5	I_1 4	D_3 3	D_2 2	D_1 1	D_o 0	P 7	C 6	I_2 5	I_1 4	D_3 3	D_2 2	D_1 1	D_o 0		
Master Space	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	MS	Blank
Upper Case	1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	1	UC	UC
Lower Case	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	LC	LC
Line Feed	0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	LF	LF
Carriage Ret.	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	CR	CR
Space	0	1	0	0	0	1	0	1	1	1	0	0	0	1	0	1	SP	SP
A	0	1	0	0	0	1	1	0	1	1	0	0	0	1	1	0	A	A
B	1	1	0	0	0	1	1	1	0	1	0	0	0	1	1	1	B	B
C	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	C	C
D	0	1	0	0	1	0	0	1	1	1	0	0	1	0	0	1	D	D
E	0	1	0	0	1	0	1	0	1	1	0	0	1	0	1	0	E	E
F	1	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	F	F
G	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	G	G
H	1	1	0	0	1	1	0	1	0	1	0	0	1	1	0	1	H	H
I	1	1	0	0	1	1	1	0	0	1	0	0	1	1	1	0	I	I
J	0	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	J	J
K	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	K	K
L	0	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1	L	L
M	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	M	M
N	1	1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	N	N
O	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	0	O	O
P	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	P	P
Q	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1	0	Q	Q
R	0	1	0	1	0	1	1	1	1	1	0	1	0	1	1	1	R	R
S	0	1	0	1	1	0	0	0	1	1	0	1	1	0	0	0	S	S
T	1	1	0	1	1	0	0	1	0	1	0	1	1	0	0	1	T	T
U	1	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	U	U
V	0	1	0	1	1	0	1	1	1	1	0	1	1	0	1	1	V	V
W	1	1	0	1	1	1	0	0	0	1	0	1	1	1	0	0	W	W
X	0	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	X	X
Y	0	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	Y	Y
Z	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	Z	Z

ALPHANUMERIC

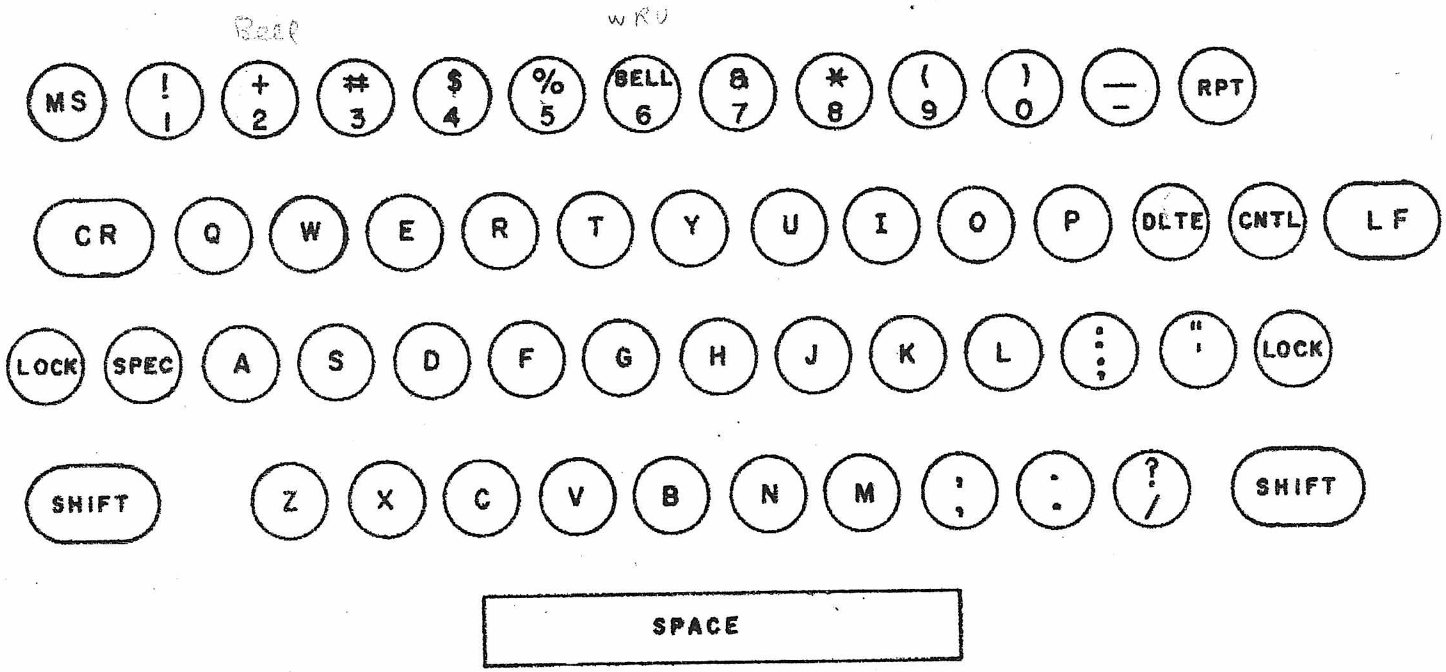
PROPOSED CHARACTER	STANDARD CODE (ODD PARITY)								PAPER TAPE (EVEN PARITY)								FIELDATA	T
	P	C	I2	I_1	D_3	D_2	D_1	D_0	P	C	I2	I_1	D_3	D_2	D_1	D_0		
	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0		
)	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	)	)
!	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	-	!
+	0	1	1	0	0	0	1	0	1	1	1	0	0	0	1	0	+	Bell
#	1	1	1	0	0	0	1	1	0	1	1	0	0	0	1	1	<	#
$	0	1	1	0	0	1	0	0	1	1	1	0	0	1	0	0	=	$
%	1	1	1	0	0	1	0	1	0	1	1	0	0	1	0	1	>	%
Bell	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	—	WRU
&	0	1	1	0	0	1	1	1	1	1	1	0	0	1	1	1	$	&
*	0	1	1	0	1	0	0	0	1	1	1	0	1	0	0	0	*	*
(	1	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	(	(
"	1	1	1	0	1	0	1	0	0	1	1	0	1	0	1	0	"	"
:	0	1	1	0	1	0	1	1	1	1	1	0	1	0	1	1	:	:
?	1	1	1	0	1	1	0	0	0	1	1	0	1	1	0	0	?	?
—	0	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	!	—
,	0	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	,	,
.	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	⊕	.
Ø	0	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	Ø	Ø
1	1	1	1	1	0	0	0	1	0	1	1	1	0	0	0	1	1	1
2	1	1	1	1	0	0	1	0	0	1	1	1	0	0	1	0	2	2
3	0	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1	3	3
4	1	1	1	1	0	1	0	0	0	1	1	1	0	1	0	0	4	4
5	0	1	1	1	0	1	0	1	1	1	1	1	0	1	0	1	5	5
6	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	6	6
7	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	7	7
8	1	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	8	8
9	0	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	9	9
'	0	1	1	1	1	0	1	0	1	1	1	1	1	0	1	0	'	'
;	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	;	;
/	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	/	/
-	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	.	-
special	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	special	stunt
Delete	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Idle	Rub-out

PROPOSED CHARACTER	STANDARD CODE (ODD PARITY)								PAPER TAPE (EVEN PARITY)								FIELDATA
	P	C	I_2	I_1	D_3	D_2	D_1	D_o	P	C	I_2	I_1	D_3	D_2	D_1	D_o	
	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0	
BLANK	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BLANK/IDLE
TST	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	TST
TCL	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	TCL
TAB	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	TAB
CONTROL C.R.	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	Control CR
CONTROL Spa.	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	Control SPA
" A	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	" A
" B	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	" B
" C	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	" C
" D	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	" D
" E	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	" E
" F	0	0	0	0	1	0	1	1	1	0	0	0	1	0	1	1	" F
" G	1	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	" G
" H	0	0	0	0	1	1	0	1	1	0	0	0	1	1	0	1	" H
" I	0	0	0	0	1	1	1	0	1	0	0	0	1	1	1	0	" I
" J	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	" J
" K	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	" K
" L	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	" L
" M	1	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	" M
" N	0	0	0	1	0	0	1	1	1	0	0	1	0	0	1	1	" N
" O	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	" O
" P	0	0	0	1	0	1	0	1	1	0	0	1	0	1	0	1	" P
" Q	0	0	0	1	0	1	1	0	1	0	0	1	0	1	1	1	" Q
" R	1	0	0	1	0	1	1	1	0	0	0	1	0	1	1	1	" R
" S	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	" S
" T	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	" T
" U	0	0	0	1	1	0	1	0	1	0	0	1	1	0	1	0	" U
" V	1	0	0	1	1	0	1	1	0	0	0	1	1	0	1	1	" V
" W	0	0	0	1	1	1	0	0	1	0	0	1	1	1	0	0	" W
" X	1	0	0	1	1	1	0	1	0	0	0	1	1	1	0	1	" X
" Y	1	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	" Y
" Z	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	" Z

CONTROL CODE

PROPOSED CHARACTER	STANDARD CODE (ODD PARITY)								PAPER TAPE (Even Parity)								FIELDATA
	P	C	I_2	I_1	D_3	D_2	D_1	D_o	P	C	I_2	I_1	D_3	D_2	D_1	D_o	
	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0	
Dial 0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	Dial 0
Dial 1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	Dial 1
Dial 2	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	" 2
Dial 3	0	0	1	0	0	0	1	1	1	0	1	0	0	0	1	1	" 3
Dial 4	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	" 4
Dial 5	0	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	" 5
Dial 6	0	0	1	0	0	1	1	0	1	0	1	0	0	1	1	0	" 6
Dial 7	1	0	1	0	0	1	1	1	0	0	1	0	0	1	1	1	" 7
Dial 8	1	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	" 8
Dial 9	0	0	1	0	1	0	0	1	1	0	1	0	1	0	0	1	" 9
SOC	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	SOC
SOB	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	1	SOB
SOD	0	0	1	0	1	1	0	0	1	0	1	0	1	1	0	0	SOD
SPARE	1	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	SPARE
SPARE	1	0	1	0	1	1	1	0	0	0	1	0	1	1	1	0	SPARE
STOP	0	0	1	0	1	1	1	1	1	0	1	0	1	1	1	1	STOP
RTT	1	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	RTT
RTR	0	0	1	1	0	0	0	1	1	0	1	1	0	0	0	1	RTR
NRR	0	0	1	1	0	0	1	0	1	0	1	1	0	0	1	0	NRR
EOBK	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	EOBK
EOB	0	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	EOB
EOF	1	0	1	1	0	1	0	1	0	0	1	1	0	1	0	1	EOF
EOC	1	0	1	1	0	1	1	0	0	0	1	1	0	1	1	0	EOC
AKR	0	0	1	1	0	1	1	1	1	0	1	1	0	1	1	1	AKR
RPB	0	0	1	1	1	0	0	0	1	0	1	1	1	0	0	0	RPB
ISN	1	0	1	1	1	0	0	1	0	0	1	1	1	0	0	1	ISN
NISN	1	0	1	1	1	0	1	0	0	0	1	1	1	0	1	0	NISN
CWF	0	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	CWF
BACKSPACE	1	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	BACKSPACE
SAC	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	SAC
SPC	0	0	1	1	1	1	1	0	1	0	1	1	1	1	1	0	SPC
IDLE	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	DELETE

FIGURE I



SPEC AND T = TAB
SPEC AND W = WRU

FOR REFERENCE ONLY

PROPOSED KEYBOARD

Proposal for [illegible]-Bit 64-Character Code Set
with Control Characters in 7-Bit Set (Note 2)

	Character	Marking Code Elements		Character	Marking Code Elements		Character	Marking Code Elements		Character	Marking Code Elements		
	Blank	---	---	Ø	---	-4-	.	---	--7	O	---	-47	
	Space	0		1	0	4	,	0	7	P	0	47	
	"	1		2	1	4	A	1	7	Q	1	47	
	#	01		3	01	4	B	01	7	R	01	47	N
	⊕	2		4	2	4	C	2	7	S	2	47	W
	%	0 2		5	0 2	4	D	0 2	7	T	0 2	47	A
	'	12		6	12	4	E	12	7	U	12	47	H
	&	012		7	012	4	F	012	7	V	012	47	
	(	---	3	8	---	34	G	---	3 7	W	---	347	3
	)	0	3	9	0	34	H	0	3 7	X	0	347	1
	;	1	3	-	1	34	I	1	3 7	Y	1	347	0
	:	01	3	+	01	34	J	01	3 7	Z	01	347	/
	WRU	2	3	=	2	34	K	2	3 7	@	2	347	1
	Bell	0 2	3	<	0 2	34	L	0 2	3 7	¢	0 2	347	
	!	12	3	>	12	34	M	12	3 7	*	12	347	
	?	012	3	/	012	34	N	012	3 7	& Rubout	012	347	
Bit Nos. of Binary Code (Note 1)	B	012	347		012	347		012	347		012	347	

Note 2: See Table 2 for Control Characters and Codes.
See Fig. 1 for possible corresponding Keyboard

Note 1: These numbers represent the order of the bits in the binary code; here they are written in reverse of the usual order, i.e. B0 is the lowest order bit and the radix value 2^0; B1 = 2^1, etc.

A.L.W. 4-21-61

-2 (12-56)

Table 2

Control Characters in 7-Bit Set to operate with 6-Bit Code for Graphics

	Character	Marking Code Elements		Character	Marking Code Elements		
0	↑	---	--67	UC	---	-467	
1		0	67	LC	0	-467	
2		1	67	LF	1	-467	
3		01	67	CR	01	-467	
4		2	67	Hor Tab	2	-467	
5		0 2	67	Vert Tab	0 2	-467	
6	Unassigned	12	67	Feedout	12	-467	
7		012	67		012	-467	
8		---	3-67		---	3467	
9		0	3-67		0	3467	
10		1	3-67		1	3467	
11		01	3-67		01	3467	
12		2	3-67		2	3467	
13		0 2	3-67		0 2	3467	
14		12	3-67	Stunt	12	3467	
15	↓	012	3-67	Rubout	012	3467	
17	Blocks →	C₂			C₃		(Note 1)

Note 1: See Table 3 for identification of Blocks C₂ and C₃ (each a 6-bit 16-character set) in the 7-bit system.
Codes in Block C₃ (at present unassigned) would be obtained by depressing the Shift key while pressing the [illegible]
[illegible]

420 LEXINGTON AVENUE NEW YORK 17, NEW YORK MURRAY HILL 3-6433

DATA PROCESSING GROUP
OFFICE EQUIPMENT MANUFACTURERS INSTITUTE

JOHN H. HOWARD DIRECTOR
HERBERT S. BRIGHT ENGINEERING DIRECTOR

MAY 4 3 2/1

Subcommittee X3-2 on Character Sets and Data Formats of
ASA X3 Sectional Committee on Data Processing Standards

MINUTES

Seventh Meeting, April 26th and 27th, 1961
Engineers Club, New York, New York

SUMMARY

A report on a meeting between some of the members of X3-2 and the Committee on Military Communications Systems Technical Standards was given and discussed. The "Basic Criteria" of character sets were discussed and a revised set (attached) was prepared for committee consideration. Several methods of expanding and contracting members of a family of sets to obtain other members of the family were presented and examined in detail. A new approach and organization was presented in outline form for consideration by the members of the committee.

I. C. Liggett, Chairman

C. E. Macon, Secretary

May 4, 1961

X3-2 Sub-Committee
on
Character Sets & Data Formats
of the
American Standards Association
X3 Sectional Committee
Sponsored by
O.E.M.I.

RECEIVED
DATA PROCESSING GROUP
O. E. M. I.
1961 MAY 25 PM 2 14

MINUTES

MEETING OF APRIL 26th AND 27th

1. MEMBERSHIP & ATTENDANCE

1.1 Members Present

John Auwaerter - Teletype Corp.

Mrs. Hiz (T. H. Bonn's alternate)-
Remington Rand

L. L. Griffin - Navy Department

I. C. Liggett - IBM

C. E. Macon - Burroughs Corporation

Roy W. Reach - Minneapolis Honeywell

H. J. Smith, Jr. - IBM

H. Tholstrup - Frieden Inc.

Allen L. Whitman - Bell Telephone Lab.

John Booth (Robert Gryb's alternate)-
AT&T Company

1.2 Members Absent

Leon Bloom - NCR

H. Kleinberg - RCA

Capt. W. J. Leubbert - West Point

Lawrence J. Schoenberg - Monroe Calculating Machine

1.3 Observers Present

Willis DeHart - Defense Communications Agency

2. MINUTES

The minutes of the April 12 and 13th meeting were distributed, read and approved.

3. INFORMAL MEETING

The informal meeting to be held concurrently with the WJCC was briefly discussed and agreed upon. The meetings will be held from 9 A.M. - 5 P.M. Monday, May 8th and Thursday, May 11th and possibly the morning of Wednesday, May 10th. The place of the meetings will be the Mayan Hotel, 3049 West 8th Street, Los Angeles, California.

4. MEETING REPORT

On 4-25-61, Messrs. Auwaerter, Griffin, Liggett and Whitman met with the Committee on Military Communication Systems Technical Standards. The purpose of this meeting was to informally introduce the Committee to the activities of X3-2 and discuss mutual problems. The result of this discussion for X3-2 was a better understanding of the Military problem. The main point, as pertains to X3-2, is the following:

> The military, for lack of an industry standard and lack of definitive action to establish such a standard, developed their own standard, Fieldata. At this point in time they have a considerable amount of hardware and software under development which incorporates this standard. An industry standard which is different from the mil standard would have to present strong advantages over the mil standard before it would gain acceptance by the military.

5. BASIC CRITERIA

The "Basic Criteria", attachment C of the Minutes for the April 12th and 13th meeting, were discussed and expanded. This revised document is included as attachment "A" for consideration by the committee at the next meeting.

6. GENERAL DISCUSSION OF THE "STANDARD FAMILY"

6.1 Expansion - Contraction

Further investigation and study of the transform of a set of one size to a set of a different size was undertaken.

Needless to say it was stipulated that the message to be transformed must be limited to characters contained in the smaller of the two sets. Consideration of characters derived from such devices as mode shifts, character pairs and the like was ruled out on the basis that the characters and codes would then not meet the criteria set forth for the standard.

Two fundamental schemes were considered. These are briefly described below:

Scheme 1 - Add a Bit

This scheme is discussed in some detail below. Some study of Figure 1 will indicate the possibilities and limitations of the scheme. Basically, B_5, B_6, B_7 & B_8 are successively added to progress from a 4 level to an 8 level code. When detailed, the set for each level would be fully delineated.

"Upward" Conversion

4 level set — $B_4B_3B_2B_1$ identify the 16 characters of the "numeric" set. This set includes 0 - 9 and some specials.

5 level set — B_5 is added,

B_5 = 1 (i.e. 1 $B_4B_3B_2B_1$) is the 4 level set.

B_5 = 0 (i.e. 0 $B_4B_3B_2B_1$) identifies the 16 additional characters of the 5 level set.

4 to 5 level conversion:
For data expressed in terms of the 4 level set, insert a 1 in the B_5 position to express the same data in the 5 level set.

5 to 4 level conversion:
The data must be expressed using only those characters available in the 4 level set. If this condition is met then "ignor" or strip off the B_5 bit (which will be 1 in all cases) and the data will be expressed in the 4 level set.

General Case

N level set — B_N is added

B_N = 0 (i.e. 0 $B_{N-1} \ldots B_1$) is the N-1 level set.

B_N = 1 (i.e. 1 $B_{N-1} \ldots B_1$) identifies the characters available in addition to those of the N-1 level set.

NOTE: Depending on the placement of the N-1 level set in the N level set 0 & 1 may be inverted. If N-1 is identified by $B_N = 0$ then the N-1 set will be located in the lower half of the N set. If N-1 is identified by $B_N = 1$ then the N-1 set will be located in the upper half of the N level set. Of course this pattern would be specified when the organization of the family of sets was decided.

Example 1

Take the number 1961 in the 4 level set

	B_4	B_3	B_2	B_1
1	0	0	0	1
9	1	0	0	1
6	0	1	1	0
1	0	0	0	1

in the 5 level set

	B_5	B_4	B_3	B_2	B_1
1	0	0	0	0	1
9	0	1	0	0	1
6	0	0	1	1	0
1	0	0	0	0	1

Note that by eliminating the B_5 digit the 5 level code becomes the 4 level code. Also by simply placing an "O" in the B_5 position the 4 level coded representation becomes the 5 level coded representation.

Example 2

Consider the identifier AX 56 in 6 level code this is

	B_6	B_5	B_4	B_3	B_2	B_1
A	1	0	0	0	0	0
X	1	1	0	1	1	1
5	0	1	0	1	0	1
6	0	1	0	1	1	0

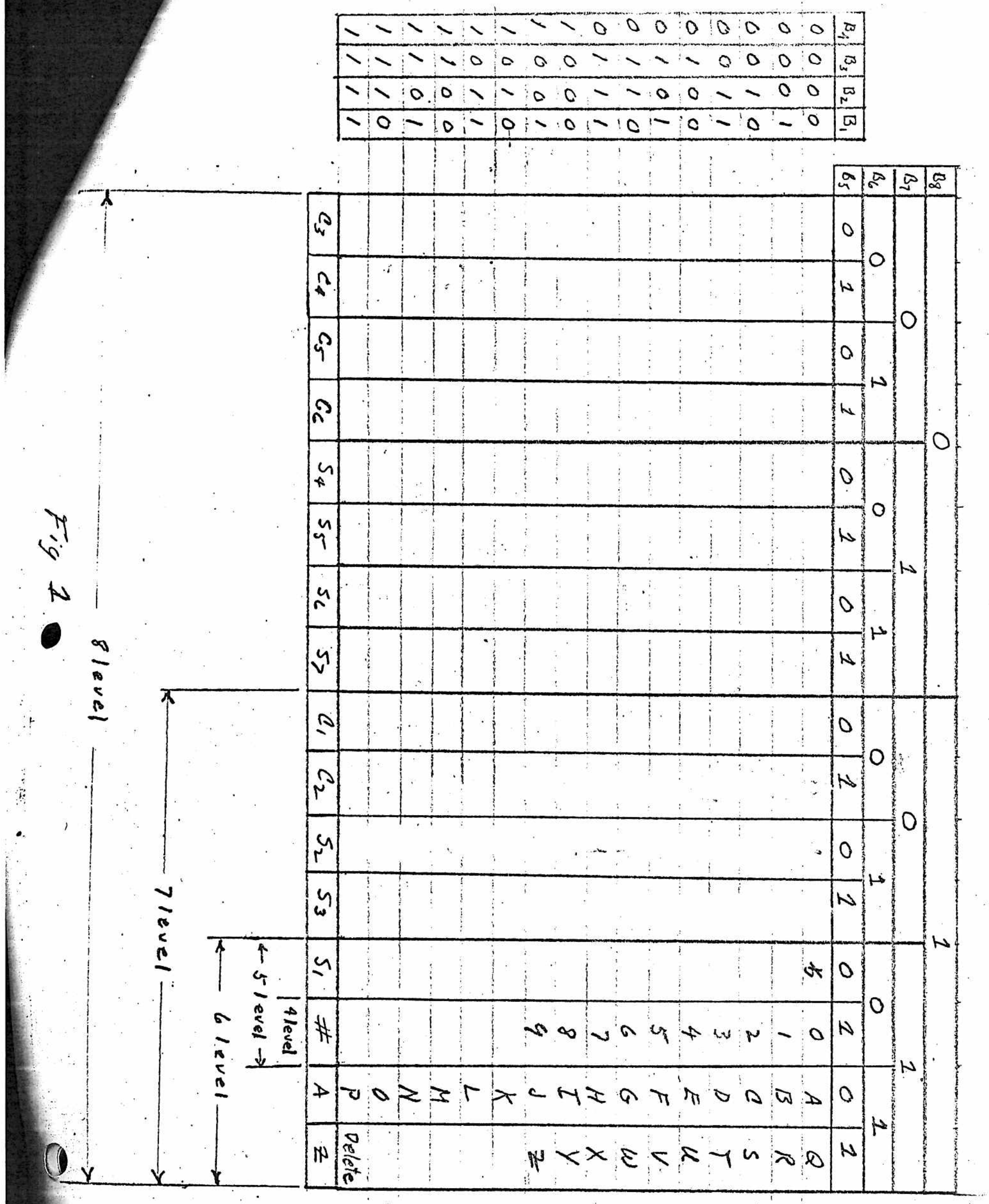


Fig 2

Z	A	#	S_1	S_3	S_2	C_2	C_1	S_7	S_6	S_5	S_4	C_6	C_5	C_4	C_3	B_1	B_2	B_3	B_4
Delete	P															1	1	1	1
	O															0	1	1	1
	N															1	0	1	1
	M															0	0	1	1
	L															1	1	0	1
	K															0	1	0	1
Z	J	9														1	0	0	1
Y	I	8														0	0	0	1
X	H	7														1	1	1	0
W	G	6														0	1	1	0
V	F	5														1	0	1	0
U	E	4														0	0	1	0
T	D	3														1	1	0	0
S	C	2														0	1	0	0
R	B	1														1	0	0	0
Q	A	0	&													0	0	0	0
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	B_5			
1		0		1		0		1		0		1		0		B_6			
1				0				1				0				B_7			
1								0								B_8			

Now if we strip off the B_6 position we have

Code						
B_5	B_4	B_3	B_2	B_1	which is	Character
0	0	0	0	0		6
1	0	1	1	1		7
1	0	1	0	1		5
1	0	1	1	0		6

The problems here, of course, is that the identifier contained characters which were available in the 6 level code but were not available in the 5 level code.

This scheme has the advantage that it is easy to understand. In terms of hardware, the transform from a larger set to a smaller set can be accomplished by simply ignoring the most significant bit position (or positions). The transform from a smaller set to a larger set can be accomplished by inserting a fixed bit (0 or 1) in the most significant bit position (or positions).

The disadvantages would be:

a) It would be impossible to maintain the "all zero's" character thru the transform (e.g. 000000 ⟶ 1000000) and, at the same time preserve the "all ones" character.

b) The basic collating sequence could not be maintained between sets.

c) It would not be possible to simply group like subsets.

Scheme 2 - Explosion

The scheme would, by means of a simple logical transform, maintain the relationship of basic types of characters. For example, the specials would be low, the numerics next, the alphabet next and the controls high. The content of each of these sections (number of characters) would vary as the level of the code. The scheme is delineated in more detail in Smith's report.

The scheme overcomes the disadvantages of Scheme 1. The transform from a larger to a smaller set could be accomplished by simply ignoring bits as before. However the transform from a smaller to a larger set would depend on the content of one (or more) of the lower order bit positions.

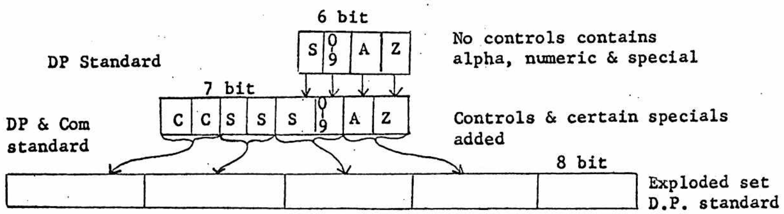
It should be noted that the expansion transform is inherently simple, both logically and in terms of hardware. Complexity is added, however, if control is included in the 6 bit (and smaller) sets.

6.2 Communications

In examining the communications problem, it appears that an 8 bit character might be reasonable as a standard. One of the bits would be a parity bit. The structure would call for 64 characters which could be printed and 64 controls and special characters (the specials in this area would not be printed by standard terminal equipment). The printable characters could be identified by examining the highest order (7th) information bit. This is essentially the Fieldata structure.

6.3 An approach

In further study of the problem the following scheme was developed. The study is not complete and a conclusion was not reached. Diagrammatically it is


6 bit
DP Standard
S
0 9
A
Z
No controls contains alpha, numeric & special
7 bit
DP & Com standard
C
C
S
S
S
0 9
A
Z
Controls & certain specials added
8 bit
Exploded set D.P. standard

(The organization of the eight bit set has not been established. Several possibilities exist all of which would not complicate the "explosion" transform).

While the details of this scheme have not been worked out, the basic philosophy is:

a) The 4, 6 and 8 bit sets are primarily Data Processing Standards.

b) Scheme 1 is employed in the transforms between 6 & 7 bit sets, Scheme 2 is used in the transform between 7 and 8 bit sets.

Note that the 8 bit set can be so structed that Scheme 1 can be employed in the 8 ⟶ 7 bit transform. The 7 ⟶ 8 bit transform will, in general, be accomplished by a Data Processor. It would seem that a processor employing an internal 8 bit code will be sufficiently large so that the slightly more complex transform of Scheme 2 will be a trivial matter to accomplish.

7. NEXT MEETING

The next regular meeting will be held June 7th and 8th in New York. Exact time and place will be announced in the forthcoming agenda.

C E Macon

C. E. Macon
Secretary

CEM:aks

DRAFT
May 4, 1961
CEM

BASIC CRITERIA FOR A STANDARD FAMILY OF CODED CHARACTER SETS

Introduction

In considering the problem of a standard which will facilitate the interchange of information between systems and between systems and associated equipment, it is implicit that such interchange must take place on a wide variety of devices which may well inherently work on basic codes of differing levels. The "standard" must be a family of coded character sets in which one member set exists for each level (i.e. 4,5,6,7, & 8 bit code levels). The interrelationship between these family members must be studied and defined so that the scheme of communication between devices of different levels is delineated and can be accomplished in the most reasonable manner.

It is understood that certain members of this family will be identified as the preferred set for information interchange and communication. Such established preference will lead to the evolution of equipment and procedures which will further facilitate the interchange and communication of information.

Reference is made to Fieldata, Section 3.6. Criteria for Alpha-numeric Code Selection and the criteria set forth in the EIA Tentative Standards Proposal for Basic Character Set Code of May 25, 1960. The 10 points set forth in Fieldata are included and all but 2 of the points of the EIA proposal. The EIA points not included are:

H) The numerals should be contiguous with the alphabet.

J) The numerics should follow the alphabet.

Specific Criteria

1) The family of sets are to facilitate the interchange of information between Systems and between Systems and Associated Equipment.

2) The codes for all symbols included in a set at a given level are to include the same number of bits.

3) Each coded set is to be structured such that expansion to a larger set or contraction to a smaller set can be achieved in a simple and logical manner.

4) Within each standard set each character and corresponding code will stand by itself and not depend on surrounding characters in the string for interpretation. That code (or codes) identified as a mode shift or escape function will meet this condition and will be considered to initiate departure from the standard set.

5) The 6, 7 & 8 level sets (i.e., 64, 128 & 256 character sets) are to include:

The alphabet
The numerals, 0-9
Certain special symbols

5.1) A useful 4 bit (16 character) numeric subset is to be defined.

5.2) The minimum control codes required to implement criteria #1 above are to be included.

5.3) For those levels that include an alphabetic subset, allowance is to be made for this subset to contain more than 26 letters. In the larger sets, allowance is to be made for the subset to include up to 32 characters. This provision will allow for incorporation of foreign alphabets.

5.4) In the larger sets, the numeric subset shall be so structured that the binary coded decimal codes for 10 & 11 (i.e., the least four significant bits) can be assigned symbols and included with the numerals. This allowance is to provide for adaptation of the code to foreign problems.

6) The numerals shall be so coded that the least four significant bits shall be the binary coded decimal form of the decimal digit which the code represents.

7) The English alphabet shall be contiguous, A-->Z, with the smallest binary number of this code area representing the letter "A".

8) The code consisting of all zeros as transmitted will be reserved for a space or blank function.

9) The all ones code will signify a delete function in paper tape.

10) Each set will contain an "Escape" character and code.

11) The simplest and most logical relationship possible shall exist between the binary sequence and the collating sequence and, wherever possible, the two shall be identical.

---

he word separator must collate low.

Special symbols commonly used in sort keys and certain punctuation symbols must collate low with respect to the alphabet and numerals.

4) The basic collating sequence (i.e., the relationship between the numeric, alphabetic and special symbol subsets) will be established giving due consideration to those sequences used in existing files and will be structured such that the existing file sequences can be dealt with.

15) The special symbols will be so placed in the set as to simplify their code generation by typewriters and similar keyboard devices.

16) The set shall be so organized that the simplest possible tests shall be adequate to distinguish and identify the basic alphabetic, numeric and special symbol subsets.

17) If possible, the set will include the COBOL character set.

18) A standard character set will be identified for general information interchange and transmission. The remaining members of the standard family are available for use in those situations in which the interchange link is known and codes of levels differing from the general information interchange and transmission standard can be advantageously employed.

19) Consideration will be given to inclusion of suitable means for data delineating and format identification in detailing the standard sets.

NMAH 310/1

FERRANTI LIMITED

PUNCHED TAPE CODES

Computer Department

Enquiries to:
London Computer Centre:
68/71 Newman Street,
London, W.1
Telephone: MUSeum 5040
and
21 Portland Place
London, W.1
Telephone: LANgham 9211

Offices, Works and
Research Laboratories:
West Gorton,
Manchester 12.
Telephone: EAST 1301

Research Laboratories:
Lily Hill,
Bracknell, Berks.

List CS 294
April 1961

PUNCHED TAPE CODES

1. INTRODUCTION

During the last two years users and manufacturers of computers, especially in Britain but also in U.S.A., have been working together to establish codes for punched paper tapes. As an outcome, some standard codes have been proposed for 4, 5, 6, 7 and 8-track tapes.

The primary aim of such work is to facilitate the interchange of tapes between users of different computers; probably this will be of most value for tapes conveying data or intermediate results, but it will also be relevant if common programming languages come into use. Secondly, this work should unify and simplify the development of associated equipment, etc.

In preparing any standard code, it is necessary to seek a proper balance between compatibility between users, and flexibility to accomodate differing requirements. Accordingly, only those characters and functions considered most important are defined, and blank places are left in each code to be used as required.

Attention has been concentrated on choosing symbols for data, rather than for programs; and it has been considered important to provide 10 and 11 as single characters for pence.

The proposed standard codes are shown in the attached tables, and are followed by certain other important tape codes.

In the tables, the tapes are viewed from the top, and are shown moving towards the user. This is the way the more important items of tape editing equipment are arranged, and it corresponds to movement from right to left in an ordinary computer input reader. The codes are arranged so that tapes may be placed in the computer input reader with the 3-track side (with respect to the sprocket holes) nearest the operator, and correspondingly in other machines.

Against each character is shown its value, obtained by regarding each combination of holes and no-holes as a binary number; in general, a hole is treated as a binary 1, and no-hole is a binary 0.

The tables are arranged in zones, each with 16 positions; the 4 less-significant bits of each character determine its position within the zone, and the remaining more-significant bits determine its zone.

There follow some comments on the various codes, drawing attention only to those points which it is thought may most interest users of Ferranti computers.

2. PROPOSED STANDARD 4-TRACK TAPE CODE

This is related to the proposed standard 5-track computer tape code. The tape has the same physical dimensions and layout as 5-track tape, with one track not used. It is primarily intended for lowest-cost equipment,

PUNCHED TAPE CODES

1. INTRODUCTION

During the last two years users and manufacturers of computers, especially in Britain but also in U.S.A., have been working together to establish codes for punched paper tapes. As an outcome, some standard codes have been proposed for 4, 5, 6, 7 and 8-track tapes.

The primary aim of such work is to facilitate the interchange of tapes between users of different computers; probably this will be of most value for tapes conveying data or intermediate results, but it will also be relevant if common programming languages come into use. Secondly, this work should unify and simplify the development of associated equipment, etc.

In preparing any standard code, it is necessary to seek a proper balance between compatibility between users, and flexibility to accomodate differing requirements. Accordingly, only those characters and functions considered most important are defined, and blank places are left in each code to be used as required.

Attention has been concentrated on choosing symbols for data, rather than for programs; and it has been considered important to provide 10 and 11 as single characters for pence.

The proposed standard codes are shown in the attached tables, and are followed by certain other important tape codes.

In the tables, the tapes are viewed from the top, and are shown moving towards the user. This is the way the more important items of tape editing equipment are arranged, and it corresponds to movement from right to left in an ordinary computer input reader. The codes are arranged so that tapes may be placed in the computer input reader with the 3-track side (with respect to the sprocket holes) nearest the operator, and correspondingly in other machines.

Against each character is shown its value, obtained by regarding each combination of holes and no-holes as a binary number; in general, a hole is treated as a binary 1, and no-hole is a binary 0.

The tables are arranged in zones, each with 16 positions; the 4 less-significant bits of each character determine its position within the zone, and the remaining more-significant bits determine its zone.

There follow some comments on the various codes, drawing attention only to those points which it is thought may most interest users of Ferranti computers.

2. PROPOSED STANDARD 4-TRACK TAPE CODE

This is related to the proposed standard 5-track computer tape code. The tape has the same physical dimensions and layout as 5-track tape, with one track not used. It is primarily intended for lowest-cost equipment, the tape perhaps being punched as a by-product of working an office machine.

3. PROPOSED STANDARD 5-TRACK COMPUTER TAPE CODE

This follows the Elliott (405) tape code, when in its form with 10 and 11. While being quite good for computer use, it has the same Figure Shift, Letter Shift, and Who Are You as the Telex code; these may be inportant for data transmission over normal telegraph circuits.

Numbers and some important symbols and functions have an odd number for holes for checking; then the 5th bit is removed leaving numbers in proper sequence (as in Pegasus/Mercury Code). The code has not, however, the optimum checking characteristics; in particular, five of the numbers can degenerate into No Action by the loss of a single hole.

Making Letter Shift and Erase the same is contrary to Ferranti practice; but it is normal for Telex and for Elliott working, and Elliotts have programming techniques and conventions to deal with it.

5-track tape is 11/16ths inch wide; somtimes the 5 tracks are punched in 7/8ths inch wide tape, the extra space (on the 2-track side) being used for printing.

It seems unlikely that Ferranti systems will use this tape code much, although it can be accepted on all Ferranti computers (except Sirius), using suitable input/output routines.

4. PROPOSED STANDARD 6-TRACK TAPE CODE

The 6 tracks are punched in tape of the same physical dimensions (7/8ths inch wide) as 7-track tape, one track not being used. The code has no inherent checking properties.

Because of these points, it is difficult to visualise this 6-track punched tape being much used in practice. However, the code itself is designed to be particularly suitable for use within computing systems, especially for commercial work. The theory lying behind it is described in "Considerations for Choosing a Character Code for Computers and Punched Tapes", Ross, The Computer Journal, Vol.3, No.1, January 1960, p.202 (reprints available).

This basic 6-bit data code is the basis of the proposed standard 7 and 8-track tape codes, and is used inside Orion and Atlas systems. It has proved suitable for a wide variety of applications, is applicable to many peripheral and associated machines, and is capable of development in many ways including data transmission.

The code has the more important symbols and functions in a single shift. In addition, the Shift Out character permits a shift out of this set of symbols into any other, and Shift In permits a return into this set of symbols. These are locking shifts.

The Escape character, which is of great importance, means that the next **single** character has a special meaning, i.e. "escaped" out of the standard code. It is non-locking shift character. One use is that the next character should be treated as a special function, or as a warning character. It may be used for ignoring a block of faulty data, or in connection with data-verification. Another application is that the next character means "shift into another case or code and stay in it", i.e. a locking-shift. Thus it may be used to obtain compatibility between many existing and future codes, and it permits virtually unlimited extension of a standard code.

ARQ means request for re-transmission, used in certain systems of data transmission, in the event of an error.

5. PROPOSED STANDARD 7-TRACK TAPE CODE

This is the basic 6-bit code laid out on 7-track tape with the addition of an odd-parity bit in track 5.

7-track tape is 7/8ths inch wide.

Probably the dominant justification for using 7-track tape at present is that some Creed tape editing equipment can be obtained, representing very good value for money. Such equipment - keyboard punches, verifiers, and tape copiers - would be suitable for routine punching of data in quantity; but such tape editing equipment, for use with Ferranti computers, does not at present incorporate a teleprinter (c.f. the Orion/Atlas Flexowriter 7-track tape code, below).

6. PROPOSED STANDARD 8-TRACK TAPE CODE

This is the basic 6-bit code laid out on 8-track tape (1 inch wide) with the addition of an even-parity bit in track 5 and a hole in track 8 to mean lower case.

In a computer with 6-bit characters, the upper and lower cases would probably be dealt with by introducing case-shift characters, the necessary conversion from the 8th track being done by input/output routines. However, in a computer with 8-bit characters, the characters from tape can be accepted directly; the purpose of making the lower-case hole contribute value $\frac{1}{2}$ inside the computer is to interleave upper and lower case letters properly for sorting, and to permit lower case to degenerate into upper case, giving the basic 6-bit code, if the least-significant bit is ignored.

It is probable that many users will wish to replace the suffix numbers in zone 1, lower case, by other symbols, e.g. programming or commercial signs; these would be too varied to warrant standardisation.

A tape editing set (BIMA) based on an IBM typewriter, and a keyboard perforator and verifier (Creed), using this 8-track tape code, are likely to become available during 1961. There is little doubt that this code will be widely exploited in the next few years.

7. INTERNATIONAL TELEGRAPH ALPHABET NO.2; TELEX 5-TRACK TAPE CODE

This has been arranged in the table in two forms; first, arranged with the characters on tape in the same sequence as the other codes; secondly, arranged in the standard CCITT sequence.

This code does not use the concept of a value being associated with each track and, therefore, no value can be given to each character (considered as a binary number).

However, each combination of holes in tape is given an identification number, as shown in the second form of the table.

In telegraphy, each character is serialised, track 1 being transmitted first and track 5 last.

8. PEGASUS/MERCURY/SIRIUS 5-TRACK TAPE CODE

Because so much tape editing and printing equipment with this code is in use, arrangements have been made for it to be read into Orion and Atlas. A table lookup procedure will be used to convert to the basic 6-bit code inside the computer.

Note that the significance of the tracks in this code are reversed from all other codes - it is unique in this respect; this was one of the first codes to be designed for computers, and it followed the track-numbering of the CCITT code.

9. PEGASUS 4-CASE FLEXOWRITER, 5-TRACK TAPE CODE

A variant of the Pegasus 5-track code which permits the full exploitation of the facilities of a Flexowriter. Intended primarily for high-quality printing of commercial results, with upper and lower-case letters, red and black, wide documents, etc.

10. ORION/ATLAS FLEXOWRITER 7-TRACK TAPE CODE

This is the primary tape used for input and output for Orion and Atlas, both for program and for data.

The code is a minor variant of the standard 7-track tape code, with upper and lower-case working, the main difference affecting the symbols ( ) %£ ? & * which have been placed in the lower-case of zone 1. On input to the computer, these symbols are transferred to zone 0, by dropping the 5th bit from track 6. And conversely on output. This arrangement overcomes the restriction of the limited number of printing keys (44) on the keyboard of Flexowriters or other typewriters.

The table shows alternative symbols for certain tape-characters. Those for preparing Orion programs are shown first in each pair; the alternates are for Atlas programs, or for a typical machine for preparing commercial data. For the latter applications some users may require special symbols, and these should preferably be allocated in conformity with the following rules:

(i) The symbols of the basic 6-bit code should be retained;

(ii) Any symbol shown in one place in the table should not be moved to another.

(iii) Any symbol not wanted may be replaced by another not already shown.

Further, some special machines have already been designed using the following code-values:- 59, 60, 61 for Swedish letters Ä, Å , Ö , or for $\frac{1}{4}$, @, $\frac{3}{4}$, respectively; and for Flexowriter functions, 8 = change readers, 9 = change colours, 10 = non-print, 11 = print-restore.

The Creed equipment used for 7-track tapes with Orion and Atlas is arranged to punch in this code, rather than the standard 7-track tape code, so that tapes will be compatible with Flexowriter tapes.

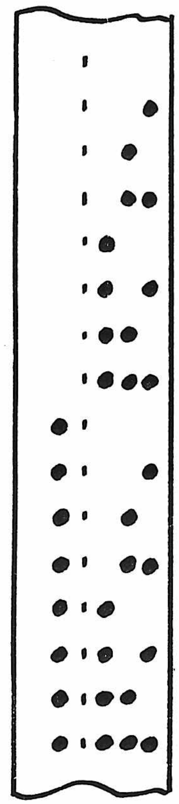
HMcGR/EL
11.4.61

© FERRANTI LTD 1961

PROPOSED STANDARD 4-TRACK TAPE CODE

TAPE MOVEMENT ↓

TOP VIEW



Track Nos.	4		3	2	1	
Values	8		4	2	1	
Value						
0		'				0
1		'			●	1
2		'		●		2
3		'		●	●	3
4		'	●			4
5		'	●		●	5
6		'	●	●		6
7		'	●	●	●	7
8	●	'				8
9	●	'			●	9
10	●	'		●		10
11	●	'		●	●	11
12	●	'	●			SPACE
13	●	'	●		●	—
14	●	'	●	●		NEWLINE
15	●	'	●	●	●	ERASE ▨

H. McG. Ross
Jan 23 1961

CS 294

PROPOSED STANDARD 5-TRACK TAPE CODE

TAPE MOVEMENT ↓

TOP VIEW

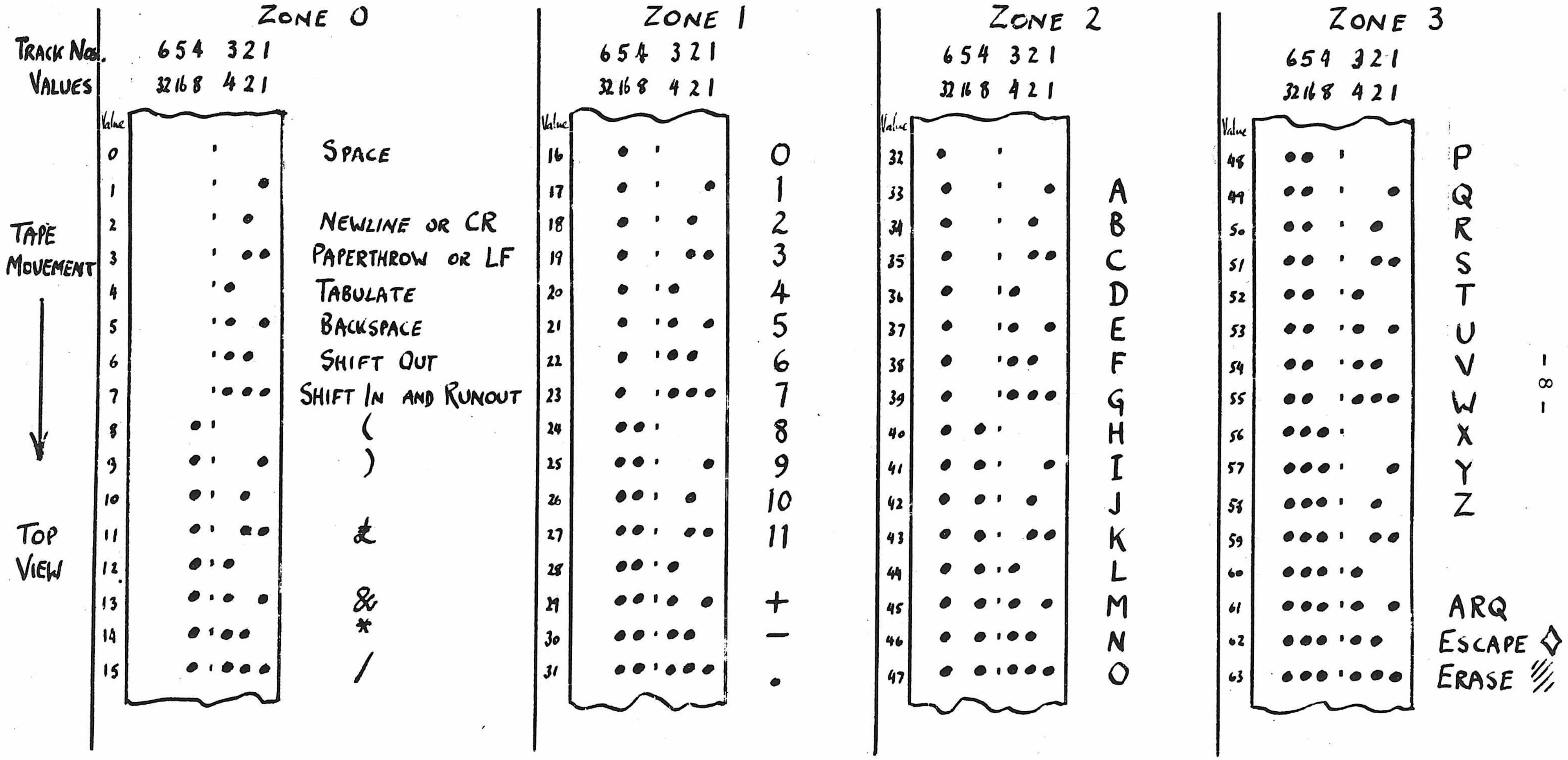
Value	5	4	3	2	1	FIGURE SHIFT	LETTER SHIFT
TRACK NOS.	5	4	3	2	1		
VALUES	16	8	4	2	1		
0						NO ACTION	
1					●	1	A
2				●		2	B
3				●	●		C
4			●			4	D
5			●		●		E
6			●	●			F
7			●	●	●	7	G
8		●				8	H
9		●			●		I
10		●		●		+	J
11		●		●	●	11	K
12		●	●				L
13		●	●		●	−	M
14		●	●	●		.	N
15		●	●	●	●		O

Value	5	4	3	2	1	FIGURE SHIFT	LETTER SHIFT
TRACK NOS.	5	4	3	2	1		
VALUES	16	8	4	2	1		
16	●					0	P
17	●				●		Q
18	●			●		WHO ARE YOU	R
19	●			●	●	3	S
20	●		●				T
21	●		●		●	5	U
22	●		●	●		6	V
23	●		●	●	●		W
24	●	●					X
25	●	●			●	9	Y
26	●	●		●		10	Z
27	●	●		●	●	FIGURE SHIFT	
28	●	●	●			SPACE	
29	●	●	●		●	CARRIAGE RETURN	
30	●	●	●	●		LINE FEED	
31	●	●	●	●	●	LETTER SHIFT/ERASE ▨	

CS 294

H. McG. Ross
Jan 23 1961

PROPOSED STANDARD 6-TRACK TAPE CODE



Value	Zone 0	Value	Zone 1	Value	Zone 2	Value	Zone 3
0	SPACE	16	0	32		48	P
1		17	1	33	A	49	Q
2	NEWLINE OR CR	18	2	34	B	50	R
3	PAPERTHROW OR LF	19	3	35	C	51	S
4	TABULATE	20	4	36	D	52	T
5	BACKSPACE	21	5	37	E	53	U
6	SHIFT OUT	22	6	38	F	54	V
7	SHIFT IN AND RUNOUT	23	7	39	G	55	W
8	(	24	8	40	H	56	X
9	)	25	9	41	I	57	Y
10		26	10	42	J	58	Z
11	£	27	11	43	K	59	
12		28		44	L	60	
13	&	29	+	45	M	61	ARQ
14	*	30	−	46	N	62	ESCAPE ◇
15	/	31	.	47	O	63	ERASE ▨

CS 294

H. McG. Ross
Jan 23 1961

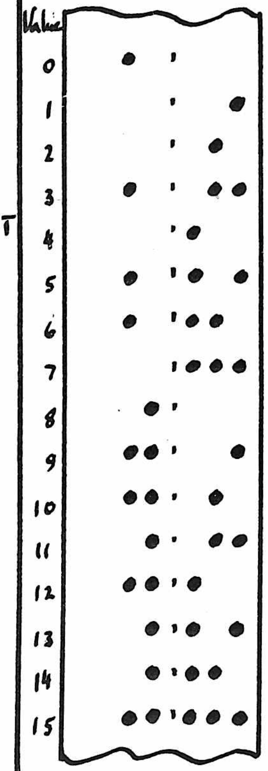
PROPOSED STANDARD 7-TRACK TAPE CODE

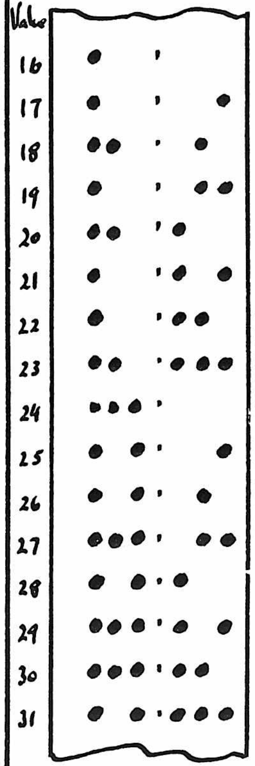
Track Nos. 7654 321

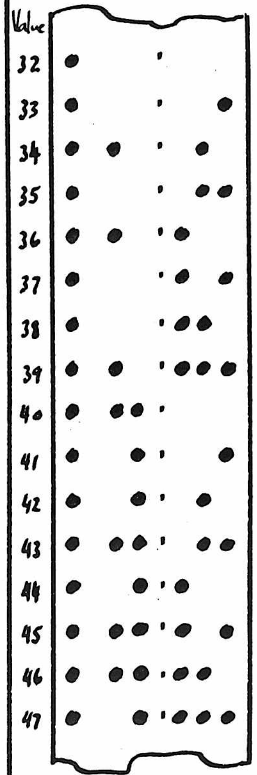
Values 32 16 P 8 4 2 1

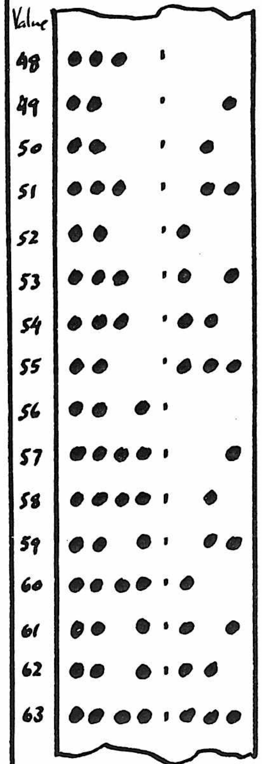
Tape Movement ↓

Top View









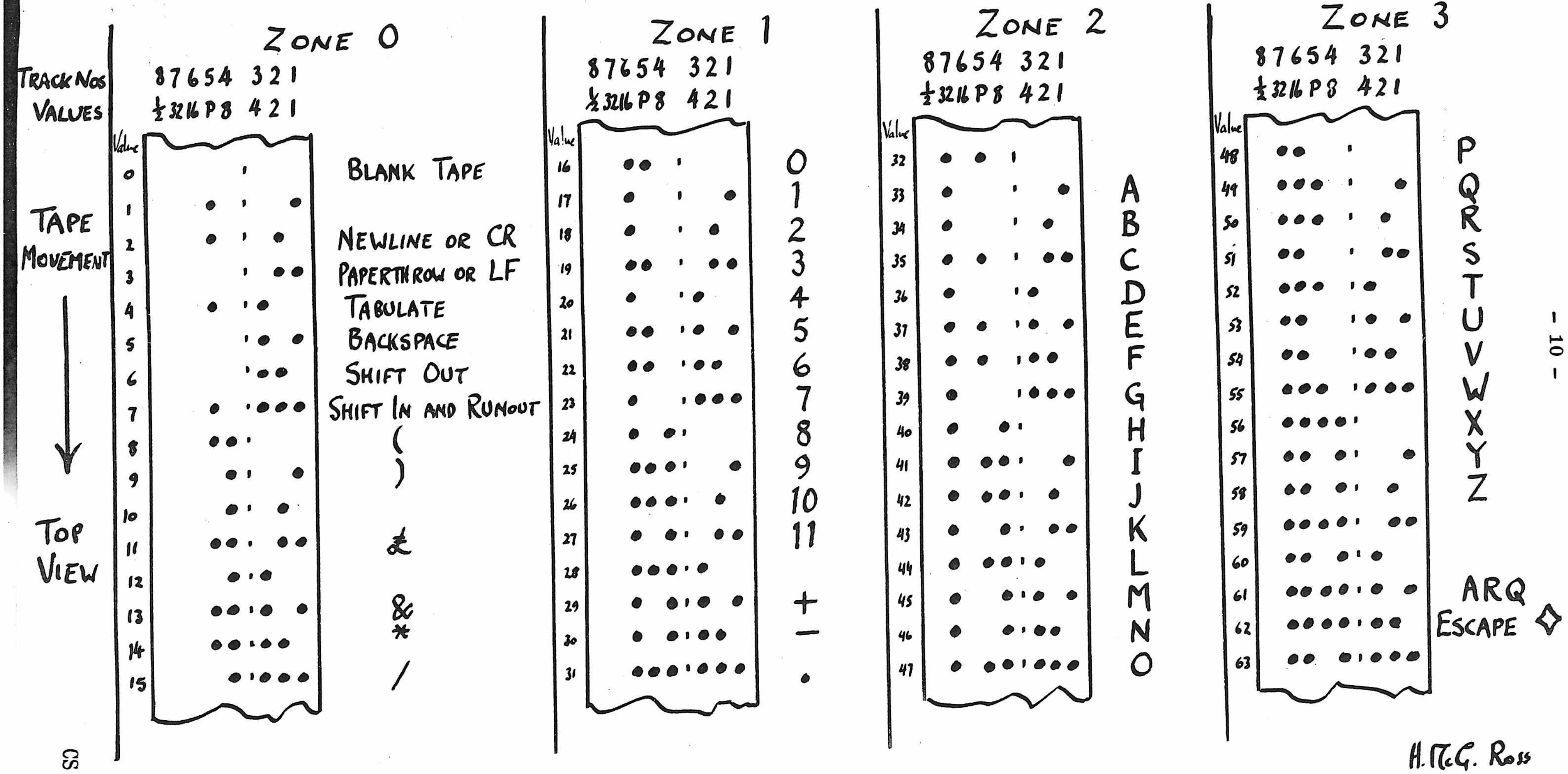
Value	Zone 0	Value	Zone 1	Value	Zone 2	Value	Zone 3
0	Space	16	0	32		48	P
1		17	1	33	A	49	Q
2	Newline or CR	18	2	34	B	50	R
3	Paperthrow or LF	19	3	35	C	51	S
4	Tabulate	20	4	36	D	52	T
5	Backspace	21	5	37	E	53	U
6	Shift Out	22	6	38	F	54	V
7	Shift In and Runout	23	7	39	G	55	W
8	(	24	8	40	H	56	X
9	)	25	9	41	I	57	Y
10		26	10	42	J	58	Z
11	£	27	11	43	K	59	
12		28		44	L	60	
13	&	29	+	45	M	61	ARQ
14	*	30	−	46	N	62	Escape ◇
15	/	31	.	47	O	63	Erase ▨

H. McG. Ross
Jan 24 1961

CS 294

PROPOSED STANDARD 8-TRACK TAPE CODE (FIRST SHEET)

UPPER CASE



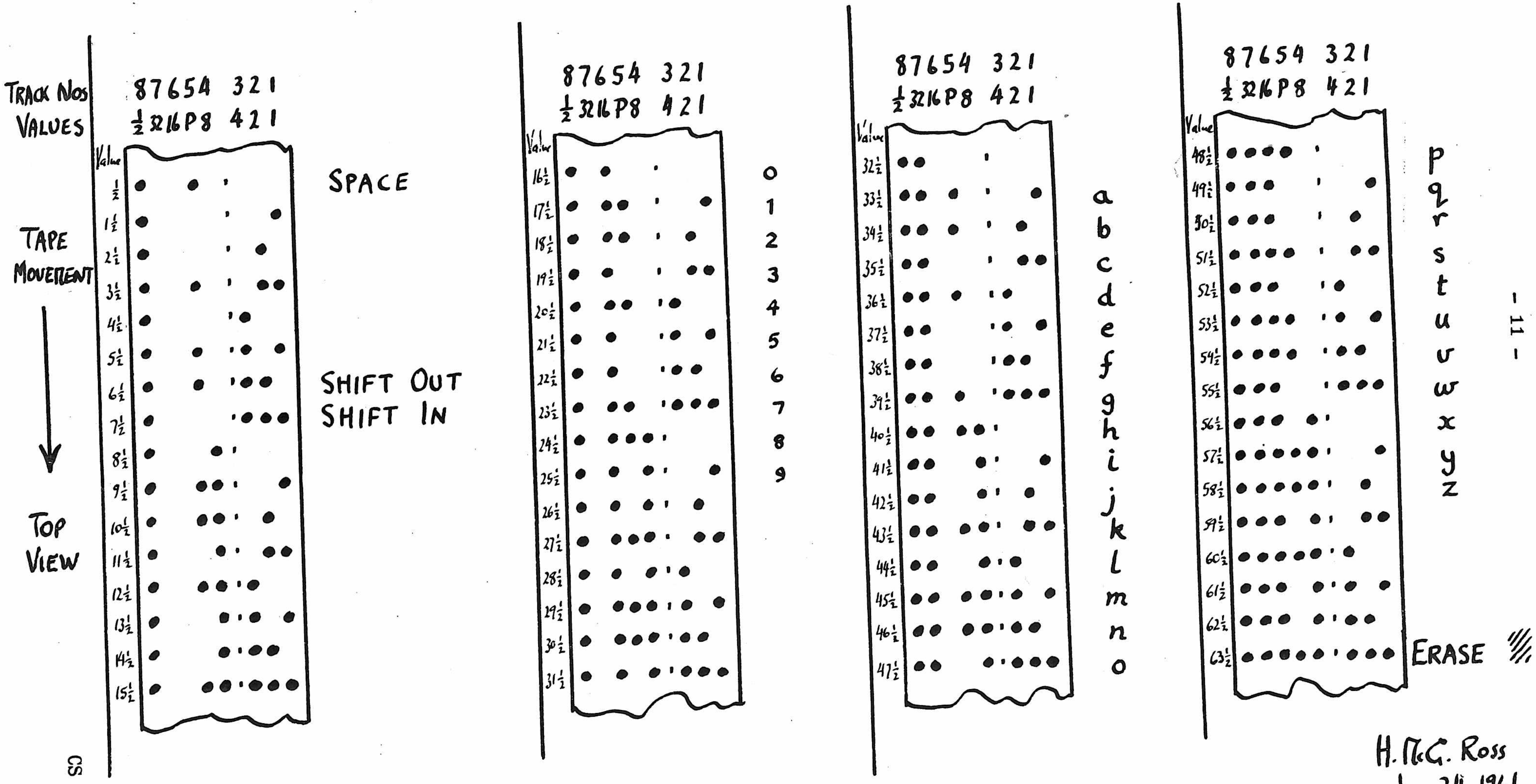
Value	Zone 0	Value	Zone 1	Value	Zone 2	Value	Zone 3
0	Blank Tape	16	0	32		48	P
1		17	1	33	A	49	Q
2	Newline or CR	18	2	34	B	50	R
3	Paperthrow or LF	19	3	35	C	51	S
4	Tabulate	20	4	36	D	52	T
5	Backspace	21	5	37	E	53	U
6	Shift Out	22	6	38	F	54	V
7	Shift In and Runout	23	7	39	G	55	W
8	(	24	8	40	H	56	X
9	)	25	9	41	I	57	Y
10		26	10	42	J	58	Z
11	£	27	11	43	K	59	
12		28		44	L	60	
13	&	29	+	45	M	61	ARQ
14	*	30	−	46	N	62	Escape ◇
15	/	31	.	47	O	63	

H. McG. Ross
Jan 24 1961

CS 29

PROPOSED STANDARD 8-TRACK TAPE CODE (SECOND SHEET)

LOWER CASE



H. McG. Ross
Jan 24 1961

CS 294

INTERNATIONAL TELEGRAPH ALPHABET No. 2; TELEX 5-TRACK TAPE CODE

ARRANGED AS OTHER COMPUTER CODES

↓ TAPE MOVEMENT

TOP VIEW

Track No. 1	2		3	4	5	Figure Shift	Letter Shift
		·				No Action	
		·			●	5	T
		·		●		Carriage Return	
		·		●	●	9	O
		·	●			Space	
		·	●		●	£	H
		·	●	●		,	N
		·	●	●	●	.	M
	●	·				Line Feed	
	●	·			●	)	L
	●	·		●		4	R
	●	·		●	●	@	G
	●	·	●			8	I
	●	·	●		●	0	P
	●	·	●	●		:	C
	●	·	●	●	●	; or =	V

1	2		3	4	5	Figure Shift	Letter Shift
●		·				3	E
●		·			●	+ or "	Z
●		·		●		Who are you	D
●		·		●	●	?	B
●		·	●			'	S
●		·	●		●	6	Y
●		·	●	●		%	F
●		·	●	●	●	/	X
●	●	·				—	A
●	●	·			●	2	W
●	●	·		●		Bell	J
●	●	·		●	●	Figure Shift	
●	●	·	●			7	U
●	●	·	●		●	1	Q
●	●	·	●	●		(	K
●	●	·	●	●	●	Letter Shift	

ARRANGED IN OFFICIAL C.C.I.T.T. SEQUENCE

Combination No.	1	2		3	4	5	Figure Shift	Letter Shift
1	●	●	·				—	A
2	●		·		●	●	?	B
3		●	·	●	●		:	C
4	●		·		●		Who are you	D
5	●		·				3	E
6	●		·	●	●		%	F
7		●	·		●	●	@	G
8			·	●		●	£	H
9		●	·	●			8	I
10	●	●	·		●		Bell	J
11	●	●	·	●	●		(	K
12		●	·			●	)	L
13			·	●	●	●	.	M
14			·	●	●		,	N
15			·		●	●	9	O
16		●	·	●		●	0	P

Combination No.	1	2		3	4	5	Figure Shift	Letter Shift
17	●	●	·	●		●	1	Q
18		●	·		●		4	R
19	●		·	●			'	S
20			·			●	5	T
21	●	●	·	●			7	U
22		●	·	●	●	●	; or =	V
23	●	●	·			●	2	W
24	●		·	●	●	●	/	X
25	●		·	●		●	6	Y
26	●		·			●	+ or "	Z
27			·		●		Carriage Return	
28		●	·				Line Feed	
29	●	●	·	●	●	●	Letter Shift	
30	●	●	·		●	●	Figure Shift	
31			·	●			Space	
32			·				Not Used	

H.R.G. Ross
Jan 27 1961

CS 294

PEGASUS/MERCURY/SIRIUS 5-TRACK TAPE CODE

TAPE MOVEMENT ↓

TOP VIEW

Track Nos.	1	2	3	4	5		
Values	1	2	4	8	16	FIGURE SHIFT	LETTER SHIFT
Value							
0						FIGURE SHIFT AND RUNOUT	
1	●					1	A
2		●				2	B
3	●	●				*	C
4			●			4	D
5	●		●			(	E
6		●	●			)	F
7	●	●	●			7	G
8				●		8	H
9	●			●		≠	I
10		●		●		=	J
11	●	●		●		—	K
12			●	●		ʊ	L
13	●		●	●		LINE FEED	M
14		●	●	●		SPACE	N
15	●	●	●	●		,	O
16					●	0	P
17	●				●	>	Q
18		●			●	≥	R
19	●	●			●	3	S
20			●		●	→	T
21	●		●		●	5	U
22		●	●		●	6	V
23	●	●	●		●	/	W
24				●	●	×	X
25	●			●	●	9	Y
26		●		●	●	+	Z
27	●	●		●	●	LETTER SHIFT	
28			●	●	●	.	.
29	●		●	●	●	n	?
30		●	●	●	●	CARRIAGE RETURN	£
31	●	●	●	●	●	ERASE	✱

H. McG. Ross
Jan 24 1961

CS 294

PEGASUS 4-CASE FLEXOWRITER, 5-TRACK TAPE CODE

TAPE MOVEMENT ↓ TOP VIEW

Track Nos.	1	2	3	4	5	FIGURE SHIFT		LETTER SHIFT	
Values	1	2	4	8	16	UPPER CASE	LOWER CASE	UPPER CASE	LOWER CASE
0						— FIGURE SHIFT AND RUNOUT —			
1	●					¼	1	A	a
2		●				½	2	B	b
3	●	●				CHANGE READERS		C	c
4			●			/	4	D	d
5	●		●			PRINT RESTORE		E	e
6		●	●			NON-PRINT		F	f
7	●	●	●			<	7	G	g
8				●		>	8	H	h
9	●			●		'	+	I	i
10		●		●		Cr	−	J	j
11	●	●		●		:	11	K	k
12			●	●		TABULATE		L	l
13	●		●	●		NEWLINE		M	m
14		●	●	●		,	.	N	n
15	●	●	●	●				O	o
16					●	£	0	P	p
17	●				●	PUNCH ON		Q	q
18		●			●	PUNCH OFF		R	r
19	●	●			●	¾	3	S	s
20			●		●	RED	BLACK	T	t
21	●		●		●	&	5	U	u
22		●	●		●	?	6	V	v
23	●	●	●		●	BACKSPACE		W	w
24				●	●	STOP		X	x
25	●			●	●	%	9	Y	y
26		●		●	●	=	10	Z	z
27	●	●		●	●	— LETTER SHIFT —			
28			●	●	●	— SPACE —			
29	●		●	●	●	— UPPER CASE —			
30		●	●	●	●	— LOWER CASE —			
31	●	●	●	●	●	— ERASE —			

H. McG. Ross
Jan 26 1961

CS 294

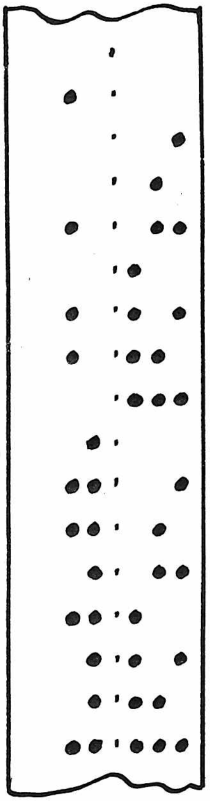
ORION/ATLAS FLEXOWRITER 7-TRACK TAPE CODE

TRACK Nos 7654 321
VALUES 32 16 P 8 4 2 1

TAPE MOVEMENT ↓

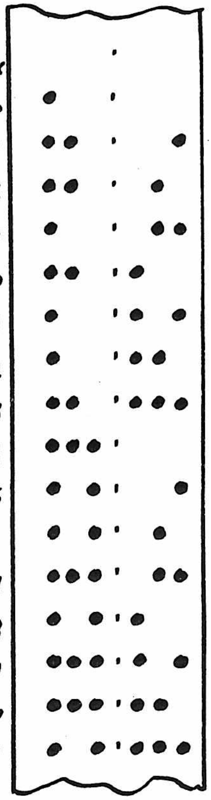
TOP VIEW

ZONE 0



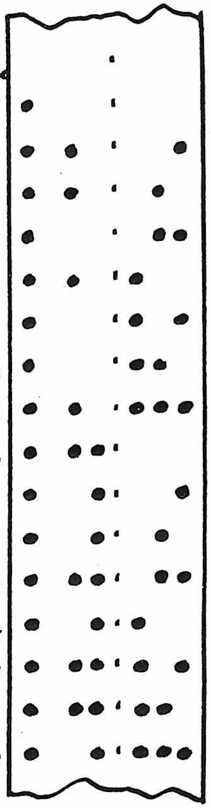
Value	UPPER CASE	LOWER CASE
0	SPACE	
1		
2	NEWLINE	
3	PAPERTHROW	
4	TABULATE	
5	BACKSPACE	
6	LOWER-CASE SHIFT	
7	UPPER-CASE SHIFT AND RUNOUT	
8		
9		
10		
11		
12	STOP	
13	PUNCH	ON
14	PUNCH	OFF
15	/	:

ZONE 1



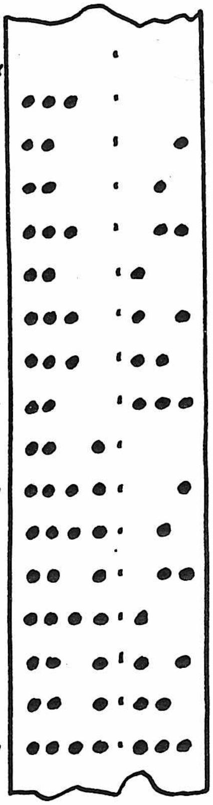
Value	UPPER CASE	LOWER CASE
16	0	'
17	1	[or ¼
18	2	] or "
19	3	< or ¾
20	4	> or @
21	5	=
22	6	—
23	7	| or ;
24	8	(
25	9	)
26	10 or α	% or $^{-1}$
27	11 or β	£ or π
28	½	?
29	+	&
30	—	*
31	•	,

ZONE 2



Value	UPPER CASE	LOWER CASE
32		
33	A	a
34	B	b
35	C	c
36	D	d
37	E	e
38	F	f
39	G	g
40	H	h
41	I	i
42	J	j
43	K	k
44	L	l
45	M	m
46	N	n
47	O	o

ZONE 3



Value	UPPER CASE	LOWER CASE
48	P	p
49	Q	q
50	R	r
51	S	s
52	T	t
53	U	u
54	V	v
55	W	w
56	X	x
57	Y	y
58	Z	z
59		
60		
61		
62	ESCAPE	
63	ERASE ▨	

H.R.G. Ross
Jan 25 1961

FRED W. SMITH, Assistant to Telegraph Equipment Engineer

A Proposed New Teleprinter Code for Error Detection

In recent years a considerable amount of study has been devoted to error-detection and error-correction systems for telegraph and data communications. A number of articles on this subject have appeared in Technical Review and many papers on the subject are presented each year before various engineering societies.

Emphasis has been placed on error detection, since a detected error can always be corrected, and also on use of systems and methods for data communication, where an undetected error can prove costly. The various codes used in telegraph and data communications have received a great deal of attention and many new codes have been developed for special purposes, with emphasis on error detection and error correction. The 5-unit Baudot code, long recognized as illogical in many respects, has also been studied[1] with a view to improving the character assignment to provide some protection against undetected errors. To date, no 5-unit code with an improved character assignment has been used in the United States, but such a code is now in limited use in Europe.

The 8-Level Data Code

In data communications there is a definite trend towards use of codes having more than five intelligence levels, or pulses, and some communications specialists feel that the 8-level code will some day be the standard code for data and message communication. One 8-level code which appears to be favored for this purpose at present would utilize five of the intelligence levels to obtain 32 possible code combinations. A sixth level would be used as a shift level; that is, to distinguish between upper-case (figures or symbols) and lower-case (letters) printed characters. A seventh level would be used for a vertical parity check; that is, a seventh marking pulse would be automatically added where necessary to insure that every combination contained an odd (or even) number of marking pulses. An eighth level would be used for control purposes and could be utilized for any purpose required. The five levels in this code used for obtaining the 32 code combinations could employ either the present Baudot code assignment or a new and more logical assignment to provide some protection against undetected errors.

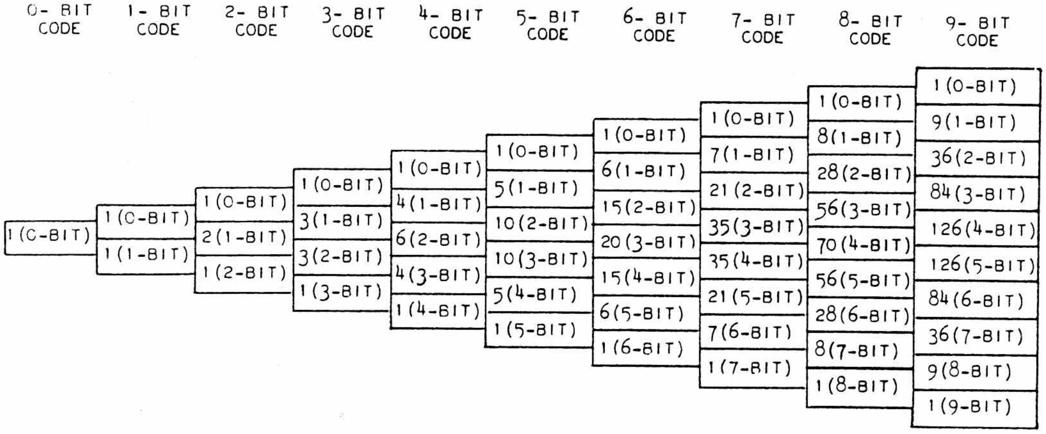
Use of an 8-level code to obtain only 32 possible code combinations for selecting printed characters or teleprinter functions appears to be highly inefficient because of the increased line frequency required for a given speed of operation and because of the more complex mechanical apparatus required to handle an 8-level code. The primary purpose of this article is to propose at least one attractive alternative to the use of an 8-level code in future telegraph systems, and to attempt to stimulate further study of a possible standard code for future telegraph and data communications use. Since adoption of a new 8-level code in the telegraph industry would involve extensive design of new equipment and/or redesign of existing equipment, a most careful and thorough consideration of all factors involved is justified in order to avoid placing undue limitations on the versatility and flexibility of any new system developed.

The principal advantage in using a sixth level as a shift level appears to be that it makes possible a keyboard more like a typewriter keyboard than any teleprinter keyboard now in use; that is, a 4-row keyboard which permits transmitting a number without first operating a figures-shift key can be designed. To send an upper-

April 1961

Vol 15 no 2

case character other than a number, however, it would be necessary simultaneously to depress a shift key and a character key in order to insert the sixth marking pulse. This might be attractive to customers who do not use the present 3-row teleprinter keyboards.



0-BIT CODE	1-BIT CODE	2-BIT CODE	3-BIT CODE	4-BIT CODE	5-BIT CODE	6-BIT CODE	7-BIT CODE	8-BIT CODE	9-BIT CODE
1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)	1 (0-BIT)
	1 (1-BIT)	2 (1-BIT)	3 (1-BIT)	4 (1-BIT)	5 (1-BIT)	6 (1-BIT)	7 (1-BIT)	8 (1-BIT)	9 (1-BIT)
		1 (2-BIT)	3 (2-BIT)	6 (2-BIT)	10 (2-BIT)	15 (2-BIT)	21 (2-BIT)	28 (2-BIT)	36 (2-BIT)
			1 (3-BIT)	4 (3-BIT)	10 (3-BIT)	20 (3-BIT)	35 (3-BIT)	56 (3-BIT)	84 (3-BIT)
				1 (4-BIT)	5 (4-BIT)	15 (4-BIT)	35 (4-BIT)	70 (4-BIT)	126 (4-BIT)
					1 (5-BIT)	6 (5-BIT)	21 (5-BIT)	56 (5-BIT)	126 (5-BIT)
						1 (6-BIT)	7 (6-BIT)	28 (6-BIT)	84 (6-BIT)
							1 (7-BIT)	8 (7-BIT)	36 (7-BIT)
								1 (8-BIT)	9 (8-BIT)
									1 (9-BIT)

Figure 1. Steeneck's analysis chart of binary codes up to nine levels.

Use of a vertical parity check for each character transmitted provides appreciable protection against occurrence of undetected errors and a vertical parity check can be made with relatively simple circuitry. However, such a check is not nearly as effective as a horizontal parity check and the use of an additional intelligence level solely for the purpose of providing a vertical parity check deserves careful study. In any event, preliminary tests indicate the latter to be only 90-percent effective.

Control characters are, of course, essential in most modern telegraph and data communications systems and it is essential to provide for such characters in any future code developed. Again, however, use of one additional intelligence level solely for this purpose should be carefully evaluated.

A Proposed New 6-Level Code

In a recent article in TECHNICAL REVIEW[1] Robert Steeneck pointed out some of the error-checking possibilities concealed within the 5-level code. Subsequent to preparation of his article, Steeneck developed a "Christmas Tree" analysis chart of all binary codes up to and including the 9-level code. Steeneck's analysis chart (Figure 1) shows, for each code, the number of combinations possible for each number of marking pulses (or bits) possible. For example, in a 2-level binary code there is one combination which contains no marking pulses, two combinations which contain one marking pulse, and one combination which contains two marking pulses. While it is not possible to have a binary code containing zero levels, Steeneck added a theoretical zero-level code for esthetic reasons.

The possibilities revealed by the Steeneck chart and the design work now being done on at least one 6-level teleprinter led to consideration of the possible use of a 6-level code with the sixth level being used for purposes other than shift control. Referring to Figure 1, it can be seen that the 6-level binary code contains 6 combinations with a single marking pulse, 20 combinations with "triples," and 6 with "quintuples." There are 15 "doubles" and 15 "quadruples." The combination with no marking pulse and the one with all 6 pulses marking complete the 64 combinations possible with a 6-level binary code.

Note that there are 32 combinations which contain an odd number of marking pulses and 32 which contain an even number, if zero is considered an even number. It is thus possible to obtain 32 combinations of the 6-level code, each of which will have an odd number of marking pulses, or 32 combinations, including the all-spacing combination, each of which will have an even number of marking pulses. This, of course, is the same number of combinations obtained with the standard 5-unit Baudot code.

Figure 2 shows all of the combinations in the 6-level code which contain an odd number of marking pulses, and Figure 3 shows the combinations which contain an even number of marking pulses. (Note that the code levels in both figures have been numbered from 0 to 5, rather than 1 to 6, to correspond to the numbering system recommended by the AIEE Committee on Tape Standardization.) Thus, either of these two groups of combinations could be used to obtain the 32 different characters now commonly used in telegraphy, and either group would provide a built-in parity check, either odd or even, depending on the group chosen. If either of these two codes is adopted for future record communications use, a serious effort should be made to make it a "common language" code, equally useful for message and data communications. Some of the factors which should be studied will be discussed in the following paragraphs in an effort to determine which of the two proposed codes appears more attractive.

Protection against errors can be provided to a considerable degree by careful assignment of the most frequently used characters to the code combinations least susceptible to errors. In making such an assignment, the following factors are most significant:

1. *A decision should be made as to which characters should be given maximum protection and an order of priority established for the 32 characters.*
2. *The many types of errors which occur in telegraph communications and their frequency of occurrence should be determined and an order of priority of the code combinations established, based on the frequency of occurrence of the most common types of errors.*
3. *The two orders of priority should then be matched to obtain the most desirable code assignment.*

Perhaps there are other factors which could be considered but these appear to be the most important.

Several years ago, Western Union made a study of the frequency of occurrence of the letters of the alphabet in the English language. Although this study was not based on telegraph messages, but rather on newspaper stories, it is unlikely that telegraph messages would be radically different. The order of occurrence of each

6 "singles"						*20 "triples"*																				*6 "quintuples"*					
0						0	0	0	0	0	0	0	0	0	0											0	0	0	0	0	
	1					1	1	1	1							1	1	1	1	1	1					1	1	1	1		1
		2				2				2	2	2				2	2	2				2	2	2		2	2	2		2	2
			3				3			3			3	3		3			3	3		3	3		3	3	3		3	3	3
				4				4			4		4		4		4		4		4	4		4	4	4		4	4	4	4
					5				5			5		5	5			5		5	5		5	5	5		5	5	5	5	5
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

Figure 2. A 6-level binary code using only combinations with an odd number of marking pulses. A number denotes a marking pulse and absence of a number denotes a spacing pulse.

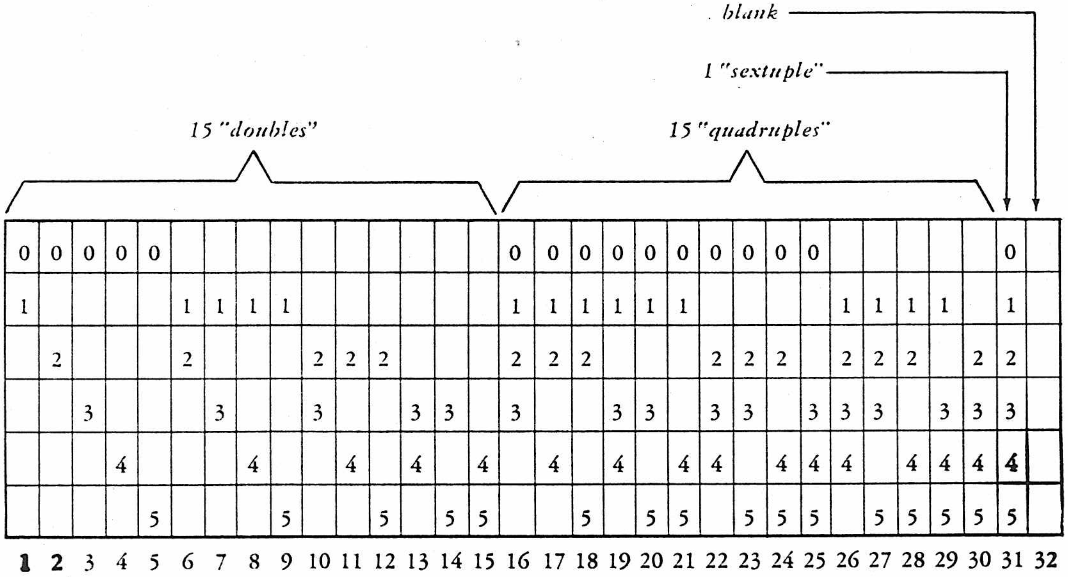
letter, with the frequency of occurrence per 1000 letters, was:

Letter	*Occurrence Per 1000 Letters*	*Letter*	*Occurrence Per 1000 Letters*
E	126	U	30
T	90	P	27
R	83	M	25
I	76	Y	21
N	76	G	18
O	74	W	14
A	72	V	13
S	58	B	11
D	40	X	5
L	36	K	3
C	33	Q	3
H	33	J	2
F	30	Z	1

Considering only the letters of the alphabet, a priority based on the foregoing tabulation could be assigned. However, in telegraph and data communications, protection of numbers is of utmost importance, since it is difficult to detect an error occurring in a number whereas an error in a letter in normal English can usually be detected and corrected. Also, protection of the five teleprinter functions (carriage return, line feed, figures shift, letters shift and space) must be considered.

It would be highly desirable to assign the ten numerals to the ten most frequently used letters in the alphabet, so that maximum protection could be given simultaneously to the numbers and to the ten most frequently used letters. Unfortunately, this would involve a radical change in the layout of 3-row keyboards. With a 4-row keyboard, the change could be made without seriously affecting the present standard layout, since the numbers keys would be separated from the letters keys and the code combinations assigned to the numbers could correspond to any group of ten letters.

Reassignment of the numbers would make it necessary to change some of the punctuation marks, but this could probably be tolerated. On the other hand, the 3-row keyboard is based on the physical relationship between the numbers and letters keys on standard typewriters and any change would make it difficult for a touch typist to use the touch system on such a keyboard. For purposes of this discussion, then, it will be assumed that



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
0	0	0	0	0											0	0	0	0	0	0	0	0	0	0						0	
1					1	1	1	1							1	1	1	1	1	1					1	1	1	1		1	
	2				2				2	2	2				2	2	2				2	2	2		2	2	2		2	2	
		3				3			3			3	3		3			3	3		3	3		3	3	3		3	3	3	
			4				4			4		4		4		4		4		4	4		4	4	4		4	4	4	4	
				5				5			5		5	5			5		5	5		5	5	5		5	5	5	5	5	

Figure 3. A 6-level binary code using only combinations with an even number of marking pulses. A number denotes a marking pulse and absence of a number denotes a spacing pulse.

3-row keyboards will continue in use in telegraph communications and that the numbers cannot be assigned to other letters; i.e., the present assignment must be maintained. The numbers one through zero are at present assigned to the letters QWERTYUIOP, in that order. The letters QWYUP are not among the ten most frequently occurring letters of the alphabet, but the remaining five in the "numbered" group are. All ten of the letters in this group must be protected, however.

An "average" word consists of five letters, so that a teleprinter must perform a spacing function an average of one time in every six operations of the teleprinter. The frequency of occurrence of the spacing character is thus approximately 200 times per 1200 characters and this frequency is even higher than that of the letter E. Although loss of, or an error in a spacing character seldom results in an undetected error, this character does deserve a high priority in the code assignment because of its high frequency of occurrence. Errors in the figures and letters shift characters seldom result in an undetected error and they are less frequently used than the space; therefore, these characters deserve less priority than the space. However, it is more difficult to correct errors due to shift failure without resending the message than it is to correct a spacing error, so that these two important function characters should be given some priority over little-used characters. Similarly, the carriage return and line feed characters seldom result in an undetected error but a single failure of one of these two characters can result in making an entire line of printing unintelligible. They thus deserve a rather high priority, perhaps higher than that assigned to the letters and figures characters.

Considering the three foregoing factors (that is, frequency of occurrence of letters, protection of numbers, and protection of functions), a number of logical orders of the priority of the 32 characters now used could be developed, depending on the weight given to each factor. The following order is suggested as just one possibility:

Priority	*Character* *L.C.*	*Character* *U.C.*
1	Space	
2	E	3
3	T	5
4	R	4
5	I	8
6	O	9
7	U	7
8	P	0
9	Y	6
10	W	2
11	Q	1
12	Car. Ret	
13	Line Feed	
14	Figures	
15	Letters	
16	N	,
17	A	–
18	S	'
19	D	$
20	L	)
21	C	:
22	H	#
23	F	¶
24	M	.
25	G	&
26	V	;
27	B	?
28	X	/
29	K	(
30	J	Bell
31	Z	"
32	Blank	

Types of Errors and Frequency

A number of studies have been made by Western Union to determine the types of errors and their frequency of occurrence in messages switched over the public message network and over one private wire network. The statistics resulting from these various studies do not agree in all respects but there is agreement as to which types of errors occur most frequently.

Loss of a single marking pulse is the most common type of nonoperator error and the field studies reveal that from 37.7 to 72 percent of all nonoperator errors are due to loss of a single pulse. The first pulse is lost more frequently than any other and the fifth pulse ranks second in frequency of loss. In normal message traffic each of these two pulses occurs in 40 percent of all characters transmitted. The third pulse ranks third in frequency of loss and occurs in 53 percent of all characters. The second pulse, which occurs in 48 percent of all characters, and the fourth pulse, which occurs in 40 percent of all characters, together account for fewer total errors due to loss of a single pulse than does the first pulse. The studies also revealed that a character containing a single marking pulse is more susceptible to error due to loss of a single pulse than a character containing several marking pulses.

Gain of a single pulse caused from 4.9 to 14.3 percent of all errors. (These results do not agree with controlled laboratory tests made on marginal equipment and circuits, which show that 90 percent of all errors[1] are due to gain or loss of a single pulse.) The remaining errors were caused by simultaneous loss and gain of pulses, loss of an entire character, gain of multiple pulses, and loss of multiple pulses, in that order of frequency.

It seems logical to assume that in a 6-level code the first and sixth pulses transmitted would be most susceptible to loss, rather than the first and fifth. If this assumption is accepted, any 6-level code adopted should assign the lowest possible priority to combinations containing both of these pulses and the next lowest to combinations containing only one of these pulses. Referring to Figure 1, it can be seen that a 4-level code contains six combinations with two marking pulses, one with four marking pulses, and one with no marking pulses. Therefore, a 6-level even-parity code will contain eight combinations with an even number of marking pulses, without either a first or sixth marking pulse, if the blank combination is included. However, use of the blank combination is considered undesirable and it is seldom used. Figure 1 also reveals that eight combinations with an odd number of marking pulses are available without using either the first or sixth pulse.

A high degree of protection against undetected errors in numbers and vowels should be provided by putting all numbers and all vowels in the same group; that is, in a group of combinations each of which contains the same number of marking pulses[1]. Since 4 of the 5 vowels have numbers assigned to the upper case, only 11 combinations in a group are necessary to accomplish this; it could be done with either the odd-parity or even-parity 6-level code. In the odd-parity code shown in Figure 2, only "triples" could be used. In the even-parity code shown in Figure 3, either "doubles" or "quadruples" could be used. In the odd-parity code there are only 4 combinations of "triples" which contain neither a first nor a sixth pulse. In the even-parity code there are 6 "doubles" which contain neither a first nor a sixth pulse and only one "quadruple" which contains neither of these two pulses. Use of the even-parity code with the numbers and vowels assigned to the "doubles" therefore offers some advantage over use of the odd-parity code. The following code assignment is therefore proposed for the even-parity code, considering all of the factors previously discussed:

Priority Number	*Combination No. (Fig. 3)*	*Code Combination*	*Character*	
			L.C.	*U.C.*
1	6	– 1 2 – – –	Space	
2	7	– 1 – 3 – –	E	3
3	8	– 1 – – 4 –	T	5
4	10	– – 2 3 – –	R	4
5	11	– – 2 – 4 –	I	8
6	13	– – – 3 4 –	O	9
7	26	– 1 2 3 4 –	Figures	
8	9	– 1 – – – 5	U	7
9	12	– – 2 – – 5	P	0
10	14	– – – 3 – 5	Y	6
11	15	– – – – 4 5	W	2
12	1	0 1 – – – –	Car. Ret.	
13	2	0 – 2 – – –	Line Feed	
14	3	0 – – 3 – –	Q	1
15	4	0 – – – 4 –	A	–
16	27	– 1 2 3 – 5	Letters	
17	28	– 1 2 – 4 5	N	,
18	29	– 1 – 3 4 5	S	'
19	30	– – 2 3 4 5	D	$
20	16	0 1 2 3 – –	L	)
21	17	0 1 2 – 4 –	C	:
22	19	0 1 – 3 4 –	H	#
23	22	0 – 2 3 4 –	F	¶
24	18	0 1 2 – – 5	M	.
25	20	0 1 – 3 – 5	G	&
26	21	0 1 – – 4 5	V	;
27	23	0 – 2 3 – 5	B	?
28	24	0 – 2 – 4 5	X	/
29	25	0 – – 3 4 5	K	(
30	31	0 1 2 3 4 5	J	Bell
31	5	0 – – – – 5	Z	"
32	32	– – – – – –	Blank	

Note that the priority previously given was modified to place all of the numbers and all of the vowels in the "doubles" group, so that loss or gain of a single pulse or of two pulses would not result in printing a wrong number or a wrong vowel. The seventh ranking combination in the order of priority was then available for a character other than a number or vowel. The figures shift was assigned to this combination because of its importance in data communications. Unfortunately, this resulted in a relatively low priority for the letters N, A and S, all three of which are among the ten most frequently occurring letters in the alphabet.

The above priority list can be divided logically into three groups. The first group consists of priorities 1 through 7, which do not contain either a zero or fifth marking pulse. The space and 6 printing characters in this group would account for 649 out of every 1200 printing and spacing characters in normal message traffic. The second group consists of priorities 8 through 23, each of which contains either a zero or fifth pulse, but not both. The printing characters in this group would account for 473 out of every 1200 printing and spacing characters. Also, 3 of the 5 function characters are in this group. The third group consists of priorities 24 through 32 and each of these combinations contains both a zero and a fifth marking pulse. The characters in this group would account for only 78 out of every 1200 printing and spacing characters and there are no function characters in this group.

Perhaps a more exhaustive analysis of the error protecting possibilities in a 6-level code would result in considerable improvement in the foregoing code assignment, but this one at least would be a tremendous improvement over the 6-level code now in limited use in the telegraph industry, where the sixth level is used only for shift-unshift, a vertical parity check is not possible, and the code assignment is more or less haphazard.

The proposed code would prevent a printing error due to loss or gain of a single pulse. The 15 "doubles" combinations would prevent a printing error due to loss of two pulses, and the 15 "quadruples" would result in printing a J (or bell in upper case) if two pulses were gained. In most cases, a J (or bell symbol) erroneously printed for another character would be an obvious error and therefore would not go undetected. All of the combinations would be protected against undetected loss or gain of more than two pulses. This code offers no protection against simultaneous loss and gain of the same number of pulses within a character combination. This could happen occasionally if the error rate was unusually high and the chances of gain and loss of pulses were about equal. In practice, however, the error rate is never that high on an operating circuit and error conditions usually consistently produce pulse losses or, more rarely, consistently produce pulse gains. Therefore, the probability of compensating errors occurring in this type of error check is quite remote.

It might appear wasteful of transmission time to use only 32 combinations out of the 64 available with a 6-level code. However, the 32 combinations which contain an odd number of marking pulses could be used for such purposes as switching control or horizontal parity checking. Switching characters or horizontal parity checking characters could be preceded by a "conditioning code" combination to bypass temporarily the vertical parity checking circuit and, if desired, to put the teleprinter in a nonprint condition. Following the switching or horizontal parity checking combinations, an end-of-switching or end-of-parity-check character could restore the circuit and the teleprinter to the normal printing condition. The 6-level code used in this manner would make considerably more efficient use of transmission facilities than the 8-level code previously described and would also make it possible to use a higher transmission speed for a given bandwidth.

Mechanical Apparatus for the 6-Level Code

Design of tape transmitters, distributors, distributor-transmitters, keyboard perforators, and nontyping reperforators for use with the proposed code would present no special problems and, in fact, many of these units are available for use with a 6-level code. They could be used with the proposed code with little or no modification. A transmitter or distributor-transmitter could be used with a parity-check circuit which would stop transmission and actuate an alarm when the reading pins read an odd number of marking pulses. A similar circuit could be used with verifying contacts or reading contacts on a reperforator to perform the same function.

At present there is no start-stop teleprinter or typing reperforator which could be readily modified to operate with the proposed code, but there appear to be no serious obstacles in the design of such units. A typebar teleprinter such as the Siemens and Halske Type 100[2] could be designed readily for six levels and it is possible that a large number of standard parts could be used in the 6-level teleprinter. A teleprinter using an aggregate-motion type of selector would be more difficult to design and it is doubtful that any of the present aggregate-motion selectors could be modified to use many of the standard parts. These comments also apply to start-stop typing reperforators.

A teleprinter with a 2-color printing mechanism such as that used on most European teleprinters[2] could be used to detect a parity-check failure by energizing a magnet to shift the red half of the ribbon to the printing position. The parity-check circuit could be arranged to lock the magnet in its operated position until manually released, so that all characters following the error would print in red. An asterisk or other symbol could be printed on receipt of the blank character to guard against undetected loss of two marking pulses in one of the "doubles" combinations. A more detailed study would no doubt reveal other possibilities. A teleprinter could be designed to space on every signal train received, whether a valid character or not. This would give excellent protection against undetected loss or gain of a single pulse, the most common cause of all errors, without the use of a parity-check circuit.

Conclusion

It is doubtful that it will ever be economical to design a communication system to provide 100-percent protection against errors, but it is economically feasible to design a reliable system approaching 100-percent accuracy. The 6-level code proposed here offers extensive possibilities. Apparatus designed for use with this code would provide a very large decrease in both detected and undetected errors, without providing parity-check circuitry, but it is not possible with the meager error-study data available accurately to predict exactly how much improvement could be expected.

New carrier systems have been developed which considerably reduce transmission errors and newer systems which will result in even greater improvement are under development. Since mechanical equipment is the source of a large percentage of errors, a new and drastically different approach to error prevention in mechanical apparatus is needed to keep pace with the improvement being made in transmission facilities. It is hoped that this article will stimulate a study of this need and perhaps lead to a new approach to the error prevention possibilities within the 6-level code. It is possible that 6-level equipment such as that proposed could be incorporated into one of the broadband switching systems now under development for data communications, with resulting simplification in the system and consequent reduction in cost.

References

1. ERROR CHECKING POSSIBILITIES CONCEALED WITHIN THE 5-UNIT CODE, ROBERT STEENECK, *Western Union Technical Review*, Vol. 14, No. 2, April 1960.
2. EUROPEAN TELEPRINTER DEVELOPMENTS, FRED W. SMITH, *Western Union Technical Review*, Vol. 14, No. 4, October 1960.

A biographical sketch of the author appears in the July 1958 issue of TECHNICAL REVIEW.

Standards

EDITOR'S NOTE

This paper describes a tentative standard code which has been proposed to the membership of the ASA X3.2 Subcommittee on Character Sets and Input/Output Media. The X3.2 Subcommittee is now examining this work in detail and, in my opinion, will probably recommend as an American Standard a code which is quite similar in principle to the one described here although it may differ in some details. Because of the urgency of this work and the very great amount of effort which has already gone into the development of this proposal, it is presented here, informally, for the information and scrutiny of the ACM membership although at proof time it had not yet been reviewed by the ACM Standards Committee for consideration as an ACM Standard. The authors and this Department will welcome your comments.

*Design of an Improved** Transmission/Data Processing Code

R. W. Bemer, H. J. Smith, Jr., F. A. Williams, Jr.
IBM Corp., White Plains, N. Y.

Historically there has been strong difference of opinion in the construction of 6-bit (64-character) data codes, based upon whether the code is to be used for communications or data processing. This paper reports on investigation of an improved code which meets transmission requirements and requires very little modification for varied data processing usage.

It has been evident from the workings of the ASA Subcommittee X3.2 that the transmission people are not as adaptable to modifications as the data processing people. This is simply a matter of inflexibility of existing semi-mechanical communications equipment compared to the general-purpose nature of electronic data processing equipment.

The major obstacle lies in the collating, or ranking, sequence of the characters of the set. It is true that a large proportion of the ordered files of today are sequenced on numeric keys alone. However, a substantial proportion of these files are ordered on keys which contain alphabetic and special punctuation characters. If a standard code changes the relative ranking of such characters the presently ordered files will all have to be fully reordered to the new sequence, a process requiring a great expenditure of machine time. Transformation of one bit representation to another is relatively simple when the sequencing property is ignored. However, one should try to guarantee that the files are still in proper order after such conversion [1].

There are three inputs to the collating problem:

(1) The most prevalent ranking in the U.S. is that established by IBM equipment, particularly the 705. In order are the blank, special characters, the alphabet, the digits. The critical point here is that the digits are higher than the alphabet, for whatever reason. The United Kingdom and certain other U.S. manufacturers (Sperry Rand and RCA) rank the digits lower than the alphabet.

(2) The desire of communications people, as first evidenced by Fieldata [2, 3, 4], is to have the 6-bit set collapsible to a 5-bit Baudot-type set with effectively the same characters. This is to utilize existing Baudot-Teletype equipment with simple modification.

(3) Certain punctuation characters, by universally accepted practice, should collate low to alphabets, digits, and other special characters. For example, the following two names would normally be ordered:

Roberts, A. B.
Robertson, X.

whereas the Fieldata code, because the comma ranks higher than the alphabet, would yield an ordering:

Robertson, X.
Roberts, A. B.

Expansion and contraction between any of the 4-, 5-, 6-, 7- and 8-bit code sets demand a certain uniformity and simplicity. Thus the alphabet should be reserved to two contiguous quadrants of the four quadrants of the 6-bit set. The choice now appears as in Figure A.

---

* Revision, 15 Mar. 1961.

May 1961 v4#5

	Quadrant			
(TYPIFIED BY)	1	2	3	4
Bendix G-20, GAMMA 60	Alphabet	Alphabet	Digits	Special
Fieldata	Alphabet	Alphabet	Special	Digits, special
IBM Stretch	Blank, special	Alphabet	Alphabet	Digits
U.K. [5]	Blank, special	Digits	Alphabet	Alphabet

FIG. A

In the opinion of the authors neither the Fieldata code nor the U.K. code meet the criterion for 5-bit Baudot-like operation completely, even though that was one of the major design requirements. A Baudot type of code is formed essentially as follows:

Letters	Control
Digits and special	

32 5-bit combinations

In any 2-mode code for paper tape, three of the control codes, DELETE, FIGURE SHIFT and LETTER SHIFT, *must* invariably be common to both shifts. DELETE must be all 1's (all punched on paper tape) and MASTER SPACE must be all 0's (unpunched tape). MASTER SPACE, BLANK, and ESCAPE preferably appear in both shifts. Such controls as LINE FEED, CARRIAGE RETURN need appear in only one shift, but operation is more complicated.

Some of these functions may be combined in a single code combination. DELETE/LETTER SHIFT is a single code in Baudot. FIGURE SHIFT is synonymous with one of the three functions possible to ESCAPE. [6]

Since DL, FS and LS must be common to both modes, Fieldata loses the Y and Z of the alphabet and the − and + characters in the collapsed 5-bit mode. This is not tolerable because some words are spelled using Y and Z. Similarly, the U.K. code loses the letters F and G and the symbols . and −. The code developed in this paper is very similar to both of these codes but removes these major flaws.

All of the criteria of the Fieldata study are used here. The full spectrum of expansion and contraction among 4-, 5-, 6-, 7- and 8-bit sets is considered in addition. Thus there are the following additional criteria and remarks:

1. A collating sequence has utility in data processing codes containing alphabets; transmission codes do not require such a sequence.
2. A collating sequence has no utility in a 4-bit set.
3. A collating sequence has utility in 5- and 6-bit sets and it is desirable that the sequence correspond to the binary representations.
4. If it is assumed that the 7- and 8-bit sets contain upper and lower case forms of the same alphabet, it is impossible to have the collating sequence match the binary representations, for the case distinction is of lesser significance than the distinction between characters with different meanings. [7, 8]
5. It is not necessary that the full 4-bit set be in 16 contiguous positions in larger sets. It is only necessary that the lowest four bit positions form the dense, unduplicated set. Other bit positions may vary. However, the digits 0-9 (10, 11) should be certainly be grouped contiguously in any set.
6. Punctuation characters have natural delimiting functions and should thus collate low to both the alphabet and digits. These include, but are not limited to:

   blank . , / − : ; ' ( )   (not in ranked order)
7. Since period and hyphen are natural delimiters, they should be placed low to both alphabets and digits. However, they often serve as radix point and minus sign (which are not delimiters) in the 4-bit numeric set. There must also be a character in this set to serve as a blank; this may or may not print in the 4-bit numeric mode. Therefore any characters of the 4-bit set which are delimiters should be in a different contiguous block than the digits, so they can serve the delimiting function in larger sets. There should be some regular transformation to append bits when expanding to larger sets.
8. All expansion and contraction from and to the various set sizes shall be blind, without knowledge of the meaning of the character assigned to any bit representation, or of contextual adjacency (with the exception of FIGURE/LETTER SHIFT control in going between 5- and 6-bit sets).
9. In all expansion and contraction, MASTER SPACE must remain all 0's and DELETE must remain all 1's. ESCAPE shall always be the second highest code, one less than DELETE; thus all bits except the low order are 1. For paper tape usage, BLANK must be different from MASTER SPACE and therefore shall have all bits 0 except the low order. This guarantees that BLANK, as the primary delimiter, collates low to all other characters. It is also the complement of ESCAPE.
10. All possible caution should be exercised in alphabetic regions to provide maximum expansion for non-English alphabets (> 26 letters).

The 8 bits are represented by B_7 through B_0, high to low order. The 6-bit transmission set will be developed first. Figure 1 shows a modified Fieldata pattern with B_5 not yet assigned, reflecting criterion 9 only. $B_5 = 0$ for Fieldata, $B_5 = 1$ for U.K.

It is now obvious that LS and FS should be opposite ES and DL, not MS and BLANK, in order to maximize the number of punctuation characters following BLANK. Since the decimal digits must have their binary representation equal to the binary value, they must be placed

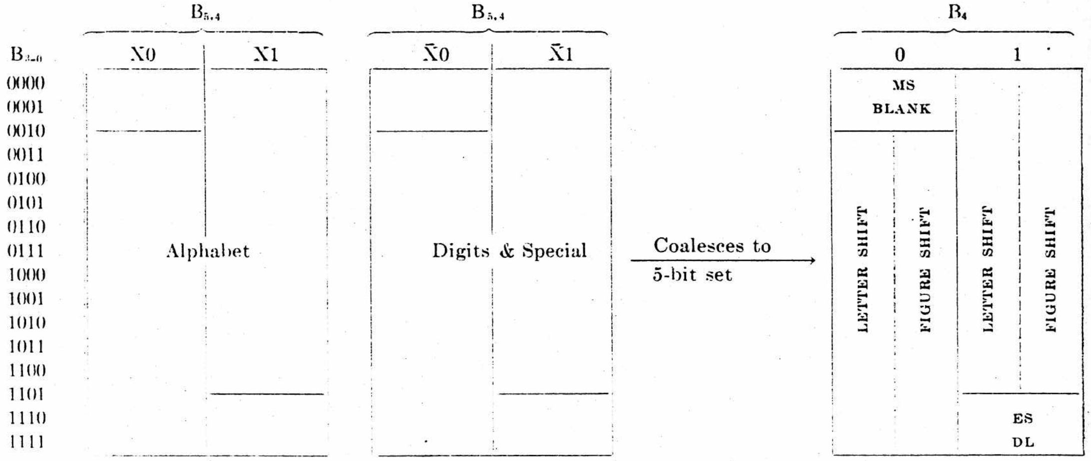


Fig. 1

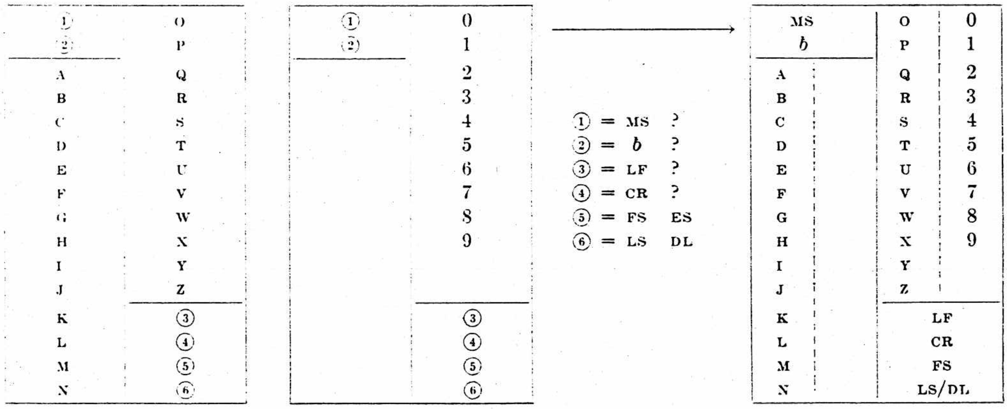


Fig. 2

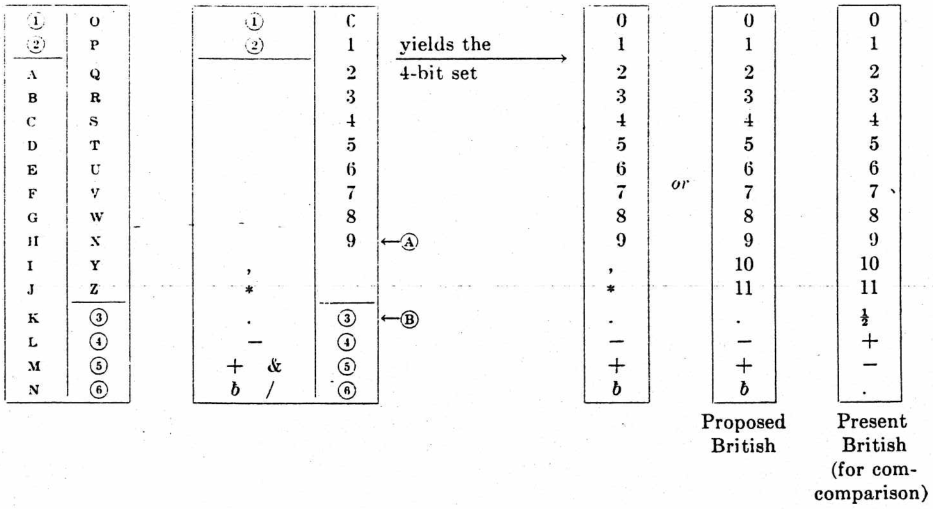


Fig. 3

For	.	−	+ &	/	,	(	)	'	:	?	=	$	# £	;	"	!	*	@	□	%	<	>
Basic CCITT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
Teletype	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓						
Basic IBM Printers	✓	✓	✓	✓	✓							✓	✓				✓	✓	✓	✓		
FORTRAN Printers	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓					✓					
1410, 7070, COBOL	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓	✓

FIG. 4

in the $\bar{X}1$ quadrant. This is shown in Figure 2. LINE FEED (LF) and CARRIAGE RETURN (CR) are necessary for character-at-a-time[1] printers associated with existing communication systems. As control signals, they are grouped with other control signals rather than with MS and BLANK, which are essentially informational.

There is space for 20 special characters in the 6-bit set, but four of these must disappear in the 5-bit set.

In conformity to most existing practice, the other six characters of the 4-bit set have been selected as:

. − +	(for self-delimiting data fields)	
ƀ	printing separator (may have graphic representation)	
,	digit grouping	most expendable for British pence (10, 11)
*	indicator for totals, etc.	

The two positions following the digit 9 are not usable for delimiters in the 6-bit set, since they will collate high. These six characters are assigned as shown in Figure 3. The pairs (.,) and (−+) are a distance of two bits apart, for easier error detection. + is used fully interchangeably with &, since & may also take the forms ⨧ ⨦ &. / has been chosen as alternate for the BLANK in the 4-bit set, since it can serve very well as a space indicator, as 00 636 505//21.

Transmission Code

B_{3-0}	$B_{5,4}$ 00	01	10	11
0000	MS	0	①	O
0001	ƀ	1	②	P
0010	"	2	A	Q
0011	$	3	B	R
0100	#	4	C	S
0101	'	5	D	T
0110	(	6	E	U
0111	)	7	F	V
1000	:	8	G	W
1001	;	9	H	X
1010	,	?	I	Y
1011	*	!	J	Z
1100	.	LF	K	③
1101	−	CR	L	④
1110	+	FS	M	ES
1111	/	LS	N	DL

$$\xrightarrow[B_4' = B_4 \forall B_5]{B_5' = B_4}$$

$$\xleftarrow[B_5 = B_4' \forall B_5']{B_4 = B_5'}$$

Data Processing Code

$B_{5,4}$ 00	01	10	11
NULL	<	O	0
ƀ	>	P	1
"	A	Q	2
$	B	R	3
#	C	S	4
'	D	T	5
(	E	U	6
)	F	V	7
:	G	W	8
;	H	X	9
,	I	Y	?
*	J	Z	!
.	K	=	CONTROL
−	L	@	
+	M	%	
/	N	□	

FIG. 5. The Proposed Standard Code

The basic set is achieved by changing the transform at Ⓐ.

$$B_4 = \bar{B}_3 \vee (\bar{B}_2 \wedge \bar{B}_1)$$

The British set *should* have 10 and 11 immediately following 9. This is achieved by changing the transform at Ⓑ.

$$B_4 = \bar{B}_3 \vee \bar{B}_2$$

Figure 4 gives the special characters specified in existing systems.

A FORTRAN-commercial substitution exists to overcome limited capacity of line printers. The correspondence is:

to = @ to ' % to (□ to)

The Bell and "who are you" functions are ignored here because they do not warrant individual characters. They are handled best by the ESCAPE mode.

B_5 may now be assigned specifically.

$X = 0$, $\bar{X} = 1$ yields a modified FIELDATA which, unless transformed, has punctuation high to the alphabet. This is not logically consistent.

$X = 1$, $\bar{X} = 0$ yields modified U.K.
" UNIVAC, MH
" RCA 501
" 704 internal

We will thus choose the latter. This choice also diminishes the number of bits or punches in numeric data, which is most frequent to data processing.

The specific proposal of Figure 5 implies either that:

(a) the data processing code is internal, and the EXCLUSIVE OR mapping takes place at the interface on reading or writing externally on media such as tape or communication lines, which utilize the transmission code, or

(b) the data processing code is merely figurative and represents the effective collating sequence obtained by a simple comparison logic in the machine.

The transmission code folds to the Baudot-like code of Figure 6. It retains all the special characters of present-day Teletype, plus the *. Although ? and !, as effective delimiters, might well precede the alphabet in the data processing code (involving a swap with < and >), to do so would remove ? and ! from the 5-bit transmission code. If the transmission people agree, this change could be considered.

As a 6-bit transmission code, ① to ④ are available. These might be used either for additional control func-

---

[1] The common term in communications is "page" printer; however, printers which print an entire page at one time must preempt this term.

MS		O	0
b		P	1
A	"	Q	2
B	$	R	3
C	#	S	4
D	'	T	5
E	(	U	6
F	)	V	7
G	:	W	8
H	;	X	9
I	,	Y	?
J	*	Z	!
K	.	LF	
L	—	CR	
M	+	FS	
N	/	LS/DL	

Fig. 6

	<	O	0
	>	P	1
	A	Q	2
$	B	R	3
#	C	S	4
'	D	T	5
(	E	U	6
)	F	V	7
:	G	W	8
;	H	X	9
,	I	Y	?
*	J	Z	!
.	K	=	
—	L	@	
+	M	%	
/	N	□	

57 + Blank

		O	0
		P	1
	A	Q	2
$	B	R	3
C #	C	S	4
F '	D	T	5
F (	E	U	6
F)	F	V	7
	G	W	8
	H	X	9
,	I	Y	
*	J	Z	
.	K	F =	
—	L	C @	
+	M	C %	
/	N	C □	

48 + Blank

Fig. 7

(Identical to Fig. 6)

MS		O	0
b		P	1
A	"	Q	2
B	$	R	3
C	#	S	4
D	'	T	5
E	(	U	6
F	)	V	7
G	:	W	8
H	;	X	9
I	,	Y	?
J	*	Z	!
K	.	LF	
L	—	CR	
M	+	FS	
N	/	LS/DL	

5-bit

$\xrightarrow{\forall}$

$\xleftarrow{\forall}$

(Identical to Fig. 7)

MS	①	O	0
b	②	P	1
"	A	Q	2
$	B	R	3
#	C	S	4
'	D	T	5
(	E	U	6
)	F	V	7
:	G	W	8
;	H	X	9
,	I	Y	?
*	J	Z	!
.	K	③	LF
—	L	④	CR
+	M	FS	ES
/	N	LS	DL

6-bit

→

←

NULL	<	O	0
b	>	P	1
"	A	Q	2
$	B	R	3
#	C	S	4
'	D	T	5
(	E	U	6
)	F	V	7
:	G	W	8
;	H	X	9
,	I	Y	?
*	J	Z	!
.	K	=	CONTROL
—	L	@	
+	M	%	
/	N	□	

Data Processing Code

Transmission Codes

Fig. 8

tions or (preferably) for additional characters of foreign alphabets. FS and LS are also available, once the complete change is made from existing equipment. Fieldata would then use these characters as UPPER CASE (UC) and LOWER CASE (LC) respectively. This provides a representation of a 7-bit code in 6-bit form, just as Baudot represents a 6-bit code in 5-bit form. Therefore the FS-UC and LS-LC correspondences are true, and either mnemonic might be used. Perhaps a new combination would be desirable, as FU (for Figure/Upper) and LL (for Letter/Lower).

= @ % and □ are placed high in the data processing code, and it is assumed they will not be used in control keys. Figure 7 shows the DP code satisfying the last set of special characters of Figure 4, plus the FORTRAN transformation and a 48-character set.

The proposed set has the special characters assigned for reasons other than matching correspondence between the digits and the characters associated with those digits on typewriter keyboards. The reasons are:

1. Some typewriters do not have keys for the digit one (1) or even for the digit (0).

2. There is no such thing as a standard typewriter keyboard in the U.S. There is a proposed British Standard, but the characters placed most uniformly, the left and right parentheses, are above 9 and 0 respectively. This conforms with much practice in the U.S., but 0 must be placed in parallel to MS in accordance with our previous rules.

3. Transmission people sometimes desire the parentheses over the 8 and 9 respectively, but this occurs only in the Luebbert revision of Fieldata (not the original Fieldata) and Ferranti computers.

4. Users normally adjust automatically to any arrangement of special characters after a day's usage.

5. Non-English keyboards differ greatly in this placement. It would be unfair to inflict one of the many English arrangements as an international standard.

6. The FORTRAN-Commercial interchange is accomplished in the proposed set by recognizing

$$B_2 \wedge \bar{B}_4 \wedge (B_3 \equiv B_5),$$

and inverting B_3 and B_5 if this condition is true. Any other arrangement greatly complicates the logical hardware necessary in converting existing printers, probably a more expensive process than converting existing Teletypewriters.

7. Any such correspondence will still require two modes of keyboard logic to generate codes.

If the transmission people could modify existing equipment with an EXCLUSIVE OR function (two relays), a completely common and collatable 6-bit code could exist, as shown in Figure 8, subject to the requirements for expansion to 7- and 8-bit sets.

REFERENCES

1. Electronic Industries Association, Basic Character Set Code, Tentative Standards Proposal 7233, May 1960.
2. LUEBBERT, W. F. Information handling and processing in large communication systems. Tech. Report 099-1, Stanford Electronics Laboratories, Stanford University, 11 July 1960.
3. U. S. Army Signal Corps, Fieldata Equipment Intercommunication Characteristics. 1 August 1959.
4. LUEBBERT, W. F. The design of the new military common-language data code. (Dittoed Copy).
5. British Standards Institution, Committee DPE 149, Draft British Standard for Punched Tape Coding, Part 1, 7 Track Code, AA (DPE) 3543, Sept. 1960.
6. BEMER, R. W. A proposal for character code compatibility. *Comm. ACM 3*, No. 2, Feb. 1960.
7. BEMER, R. W. On the design of extended character sets. 26 Jan. 1961 (Unpublished).
8. SMITH, H. F. JR. AND WILLIAMS, F. A. JR. Remarks on collating sequences. 22 Sept. 1960 (unpublished).

APPENDIX

On the Relative Position of the Alphabet, Numbers and Special Characters in a Code Set Based upon Transmission and Data Processing Characteristics

Consider a 4-quadrant arrangement of 16 states per quadrant in which Q1, Q2, Q3 and Q4 represent the octal codes 00-17, 20-37, 40-57 and 60-77. Consider also four classes of information symbols which may be placed in these quadrants:

D representing the 10 or 12 digits of the decimal or duodecimal system
S representing the class of special characters
A representing a section of the alphabet beginning with the letter A
Z representing a section of the alphabet ending with the letter Z

If each of these four classes of information is assumed to consist of up to 16 codes, they may be assigned to the 4 quadrants in any of 24 combinations. We now consider these combinations in light of their desirability for data processing and data transmission.

Transmission Considerations

1. Concepts of MASTER SPACE and BLANK are distinct. The term "BLANK" refers to the element of information used to separate words on a printed page. MASTER SPACE occupies the zeroth position.
2. The concept of ERASE or DELETE is represented by the Nth character.
3. MASTER SPACE, BLANK and DELETE are concepts required in all alphabetic or alphanumeric sets. Further MS and DL always occupy the same relative position in each set.
4. The 64-state code set is to be representable by a 32-state code using a shift mode. In this compressed representation the alphabet is to form one shift and the numbers and special characters the other.

These criteria imply:

01. A and Z cannot fold upon each other (4).
02. Q1 and Q4 cannot fold upon each other (1, 2, 3).
03. MASTER SPACE must be in Q1. (1)

Data Processing Considerations

5. The digits are represented by their pure, natural binary equivalents. Since only four bit positions are necessary to represent up to 16 states, any additional bit position in a given set must contain the same pattern of 1's and 0's for each digit.
6. No symbol other than MASTER SPACE ranks lower than BLANK in the collating sequence.
7. The alphabet is dense in collation.
8. Certain field-separating symbols including BLANK must rank lower than the alphabet in collating.

These criteria imply:

04. D cannot be in the quadrant which contains or folds on MASTER SPACE. Otherwise 0 and MASTER SPACE would become identical in some code set. (5)
05. Z cannot occupy Q1. (7)
06. MASTER SPACE and BLANK must appear as adjacent characters in the same quadrant. Otherwise some symbols will be less than BLANK or MS will have a rank higher than BLANK. (6)
07. BLANK cannot be associated with A. Otherwise some special symbols would either be lower in collating than BLANK, or field-breaking symbols would be higher in collating than the alphabet which they are intended to separate. (6, 7)
08. (7) and (8) immediately above imply that A cannot occupy Q1 since MS must be located here.

Application of Rules to the Combinations of D, A, Z, and S on Q1, Q2, Q3, and Q4

The requirements 04, 08 and 05 remove from consideration the 18 combinations beginning with D, A and Z respectively. In addition any combination in which A does not precede Z can result in a non-dense alphabet.

(*please turn to page 225*)

TABLE 1

Deviates of the Normal Function in Octal Corresponding to the Cumulative Area From .5 to 1.0

Area (Octal) Scaled 2^8	Normal Deviate (Octal) Scaled 2^{28}	Area (Octal) Scaled 2^8	Normal Deviate (Octal) Scaled 2^{28}
200	00 00000 00000.	300	00 12625 32704.
	00 00120 15457.		00 12772 42473.
	00 00240 33636.		00 13140 41660.
	00 00360 53002.		00 13307 32316.
	00 00500 73431.		00 13457 16144.
	00 00621 16043.		00 13627 77435.
	00 00741 43141.		00 14001 60730.
	00 01061 73214.		00 14154 44377.
210	00 01202 26537.	310	00 14330 34471.
	00 01322 66042.		00 14505 34031.
	00 01443 32032.		00 14663 45573.
	00 01564 03006.		00 15042 74314.
	00 01704 61253.		00 15223 42707.
	00 02025 45515.		00 15405 34463.
	00 02146 40500.		00 15570 55071.
	00 02267 42513.		00 15755 27637.
220	00 02410 54064.	320	00 16143 40227.
	00 02531 75526.		00 16333 12607.
	00 02653 30005.		00 16524 33524.
	00 02774 73427.		00 16717 27240.
	00 03116 50535.		00 17114 02241.
	00 03240 40077.		00 17312 41757.
	00 03362 42443.		00 17512 73747.
	00 03504 60346.		00 17715 25671.
230	00 03627 12161.	330	00 20121 65564.
	00 03751 60466.		00 20330 42335.
	00 04074 44256.		00 20541 43267.
	00 04217 45711.		00 20755 00012.
	00 04342 65610.		00 21173 00365.
	00 04466 24554.		00 21413 55747.
	00 04612 03410.		00 21637 22316.
	00 04736 02547.		00 22065 70106.
240	00 05062 22626.	340	00 22317 52300.
	00 05206 64474.		00 22554 65705.
	00 05333 51003.		00 23015 50463.
	00 05460 60415.		00 23262 21226.
	00 05606 13615.		00 23533 00141.
	00 05733 73512.		00 24010 07632.
	00 06062 01012.		00 24271 74500.
	00 06210 34430.		00 24560 64671.
250	00 06337 16705.	350	00 25055 11231.
	00 06466 30775.		00 25357 26315.
	00 06615 73671.		00 25667 74207.
	00 06745 70150.		00 26207 36451.
	00 07076 16612.		00 26536 46016.
	00 07227 00465.		00 27076 02014.
	00 07360 16636.		00 27446 51276.
	00 07511 72144.		00 30031 33024.
260	00 07644 03467.	360	00 30427 40773.
	00 07776 54160.		00 31042 23633.
	00 10131 65174.		00 31473 35502.
	00 10265 37466.		00 32144 76235.
	00 10421 54234.		00 32641 23053.
	00 10556 34507.		00 33363 41636.
	00 10713 61765.		00 34137 46520.
	00 11051 55313.		00 34752 51075.
270	00 11210 20025.	370	00 35633 37754.
	00 11347 33317.		00 36573 54572.
	00 11507 21011.		00 37631 00460.
	00 11647 62131.		00 41010 33357.
	00 12011 00143.		00 42354 42517.
	00 12152 74634.		00 44204 72526.
	00 12315 51606.		00 46534 51133.
	00 12461 10461.		00 52437 23555.

STANDARDS—Continued from page 217:

As can be seen, this is not the only method of applying the stated rules. Despite the order taken the rules reduce to three the number of acceptable combinations for data processing and transmission. These combinations are S, A, Z, D which is followed by IBM and S, D, A, Z which is advocated in the United Kingdom. Also possible is S, A, D, Z.

The FIELDATA arrangement A, Z, S, D is not acceptable. First, BLANK is associated with A, which means the delimiting special characters will collate higher than the alphabet. Second, if BLANK is not associated with MASTER SPACE in Q1 but is in the second position of Q3, one symbol (the character in the first position of Q3) other than MASTER SPACE must have a rank less than BLANK.

This examination of the possible combinations of S, D, A, Z merely indicates which arrangements should be further investigated with view of their expansion characteristics in sets of more than 6 bits. The analysis is intended to remove much of the confusion which has existed as to what combinations are possible and desirable for expansion and contraction. Thus, the arrangements S, A, Z, D and S, D, A, Z will be given further analysis and a choice between them made on the facility of their expansion and contraction characteristics.

Folding Considerations

At this point the methods of folding must be considered. The four quadrants may be folded in one of the two ways by the removal of either B5 or B4.

Q1 on Q2 and Q3 on Q4
Q1 on Q3 and Q2 on Q4

When the folding consideration is applied to the remaining combinations of S, A, D, Z we find,

	Q1 on Q2	Q1 on Q3
SDAZ	Note 1	Possible
SAZD	Note 2	Note 4
SADZ	Note 3	Note 1

NOTE 1. The digits cannot fold on MASTER SPACE.

NOTE 2. The transformation characteristic between 5 and 6 bits is dependent upon the combination being treated as well as the shift.

NOTE 3. This is possible by treating B_4 of the 6-bit representation as the mode bit. However, this leads to a non-dense alphabet.

NOTE 4. Z cannot fold on MASTER SPACE.

Conclusion

From the consideration of data transmission and data processing criteria we are led to a code organization of S, D, A, Z. This organization, however, should not be considered as giving the collating sequence.